IN-DEPTH BIOINFORMATICS ANALYSIS OF THE PHOSPHOPROTEOME OF TRIPLE NEGATIVE BREAST CANCER TREATED WITH A TUMOR SELECTIVE NQO1 BIOACTIVATABLE DRUG

Gitanjali Roy

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Master of Science
in the Department of Biochemistry and Molecular Biology,
Indiana University

January 2021

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Master's Thesis Committee

_____
Amber L. Mosley, Ph.D., Chair

_____
Millie M. Georgiadis, Ph.D.

_____
Mark Goebl, Ph.D.

DEDICATION

Dedicated to my parents, Dipti Roy and Ardhendu K Roy, who provided me with everything I needed. They always believed in educating their child and worked tirelessly for my education and overall growth in every aspect of my life. I couldn't have done this without their efforts, encouragement and their blessings.

I would also like to dedicate this thesis to all my teachers from school, undergraduate and graduate studies across all institutions who have contributed their knowledge and shaped my understanding of the world and science. A big thank you to all the wonderful teachers and the noble work they do in educating young minds.

And my husband, Dr. Amiya K Maji, for also supporting me throughout my journey at Indiana University and providing every possible help I needed. I couldn't have asked for a better partner in life.

ACKNOWLEDGEMENT

I would like to acknowledge my mentor Dr. Amber L. Mosley. I will be grateful for the mentorship, support and the belief in me. I started with limited experience in my area of research, but with her guidance, passion and her effort to make students competitive scientist, I gained confidence in the necessary skills to continue my journey in the biomedical field. Thank you for the collaborative effort that you established in the lab, the proteomics core, and for continuously striving to make us better scientists.

I am thankful to my committee members, Dr. Millie M. Georgiadis and Dr. Mark G. Goebl for guiding me with my thesis writing and encouraging me to think scientifically.

I am very grateful to Dr. David A. Boothman for the opportunity to collaborate with his lab and work on a novel anti-cancer agent.

I am also thankful to members and colleagues in lab, Whitney Smith Kinnaman, Dr. Sarah Peck Justice, Katlyn Hughes Burris, Dominique Baldwin, Neil McCracken and Sagara Wijeratne for their help and guidance throughout my thesis experience.

Special thanks go to the members of the Proteomics Core at Indiana University School of Medicine: - Dr. Emma Doud, Dr. Aruna Wijetarne and Guihong Qi for the continuous support and guidance during my research journey.

Gitanjali Roy

IN-DEPTH BIOINFORMATICS ANALYSIS OF THE PHOSPHOPROTEOME OF TRIPLE NEGATIVE BREAST CANCER TREATED WITH A TUMOR SELECTIVE NQO1 BIOACTIVATABLE DRUG

The main focus of this study is to elucidate changes in the proteome of triple negative breast cancer cells in response to a novel bioactivated anti-cancer agent IB-DNQ (isobutyl-deoxyniboquinone). NQO1 or NADPH:quinone oxidoreductase-1 is a detoxifying enzyme overexpressed in many solid tumors and low expression in normal cells. IB-DNQ is bio-activated by NQO1 enzyme via a futile redox cycling, producing large amounts of reactive oxygen species (ROS) in the process, which causes DNA lesions in cancer cells. The status of NQO1 is important for the IB-DNQ mediated cancer cell death.

IB-DNQ mediated therapy has great potency in killing breast cancer cells compared to PARP inhibitor Rucaparib. From this proteomics study, large changes in phosphorylation are observed in utilizing a combination therapy with low dose of IB-DNQ and PARP inhibitor Rucaparib. Protein phosphorylation events within the transcription machinery and DNA damage repair pathways are changed upon drug treatment.

Computational and bioinformatic analysis of kinases involved through kinase substrate enrichment analysis revealed changes in downstream signaling of cell cycle checkpoint proteins. CDK1 and CDK2 substrate phosphorylation was decreased in response to combination drug therapy. Based on the differential kinase activity as determined by substrate abundance, we hypothesize that since CDK1/2 plays an important role in DNA damage repair via the homologous recombination pathway, its

downregulation further abrogated double stranded break repair in BRCA deficient cells creating a state of “BRCAness”, leading to heightened sensitivity to the PARP inhibitor Rucaparib. These studies give insight into the mechanism of IB-DNQ action as an anticancer agent.

Amber L. Mosley, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PICTURES

LIST OF ABBREVIATIONS

NQO1 -	NAD(P)H Quinone Dehydrogenase 1
IB-DNQ -	Isobutyl-deoxynyboquinone
PARP -	Poly (ADP-ribose) polymerase
CDK –	Cyclin-dependent kinase
BRCA1/2 –	BReast CAncer susceptibility gene 1 or 2
FAD -	Flavin adenine dinucleotide
GO –	Gene Ontology
GSEA –	Gene Set Enrichment Analysis
KSEA –	Kinase Substrate Enrichment Analysis
SWI-SNF –	SWItch/Sucrose Non-Fermentable
POLR2A –	RNA Polymerase II Subunit A
IWS1 –	Interacts with SUPT6H, CTD Assembly Factor 1
XRCC1 –	X-Ray Repair Cross Complementing 1
TP53BP1 –	Tumor suppressor p53 binding protein 1
KAT7 –	Lysine Acetyltransferase 7
ATR –	ATM and Rad3-related protein
XPC –	DNA repair protein complementing XP-C cells
H2AX –	Histone 2A variant X
HNRNPU –	Heterogeneous nuclear ribonucleoprotein U
PSMD4 -	26S proteasome non-ATPase regulatory subunit 4
PML –	PML Nuclear Body Scaffold
SP100 –	Nuclear autoantigen Sp-100

INO80B – INO80 complex subunit B

HDAC2 – Histone Deacetylase 2

PBRM1 – Protein polybromo-1

RSF1 – Remodeling and Spacing Factor 1

SMARCC2 – SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin Subfamily C Member 2

CHD3 – Chromodomain-helicase-DNA-binding protein 3

ARID1A – AT-rich interactive domain-containing protein 1A

SLC9A1 – Solute Carrier Family 9 Member A1

MMS19 – MMS19 nucleotide excision repair protein homolog

FOXK1 – Forkhead box protein K1

ATM – Ataxia telangiectasia mutated

CHD4 – Chromodomain-helicase-DNA-binding protein 4

GATAD2B – GATA Zinc Finger Domain Containing 2B

TSC2 – Tuberin

LARP – La-related protein 1

CTTN – Src substrate cortactin

TJP1 – Tight junction protein ZO-1

ASAP1 – Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 1

SVIL – Supervillin

ABL1 – Tyrosine-protein kinase ABL1

RASAL2 – Ras GTPase-activating protein nGAP

ATAD2 – ATPase family AAA domain-containing protein 2

KDM2B – Lysine-specific demethylase 2B

MTORC1 – Mammalian target of rapamycin complex 1

MYC – Myc proto-oncogene protein

E2F – Transcription factor E2F1

CLK1 – CDC Like Kinase 1

NEK2 – Serine/threonine-protein kinase Nek2

IKBKB – Inhibitor of nuclear factor kappa-B kinase subunit beta

PRKCI – Protein kinase C iota type

MAP2K5 – Dual specificity mitogen-activated protein kinase kinase 5

CHUK – Inhibitor of nuclear factor kappa-B kinase subunit alpha

RPS6KA2 – Ribosomal protein S6 kinase alpha-2

TGFBR2 – TGF-beta receptor type-2

PDK1 – Pyruvate Dehydrogenase Kinase 1

CHEK1 – Checkpoint Kinase 1

PAK2 – P21 (RAC1) Activated Kinase 2

PLK1 – Polo Like Kinase 1

ROCK1 – Rho-associated protein kinase 1

AURKC – Aurora Kinase C

CCNB1 - G2/mitotic-specific cyclin-B1

CDC25C - M-phase inducer phosphatase 3

FOXM1 - Forkhead box protein M1

CENPU - Centromere protein U

PKMYT1 - Membrane-associated tyrosine- and threonine-specific cdc2-inhibitory kinase

PPP1R12A – Protein phosphatase 1 regulatory subunit 12A

WEE1 - Wee1-like protein kinase

MRN - Mre11, Rad50 and Nbs1

CtIP - C-terminal binding protein–interacting protein

PIN1 - Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1

RPA - Replication protein A 32 kDa subunit

Exo1 - Exonuclease 1

DNA2 - DNA replication ATP-dependent helicase/nuclease DNA2

AGC - Protein kinase A, G, and C families (PKA, PKC, PKG)

CMGC - Kinase family containing cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAP kinases), glycogen synthase kinases (GSK) and CDK-like kinases

CK1 - Casein kinase 1 family

TK - Tyrosine Kinase

INTRODUCTION

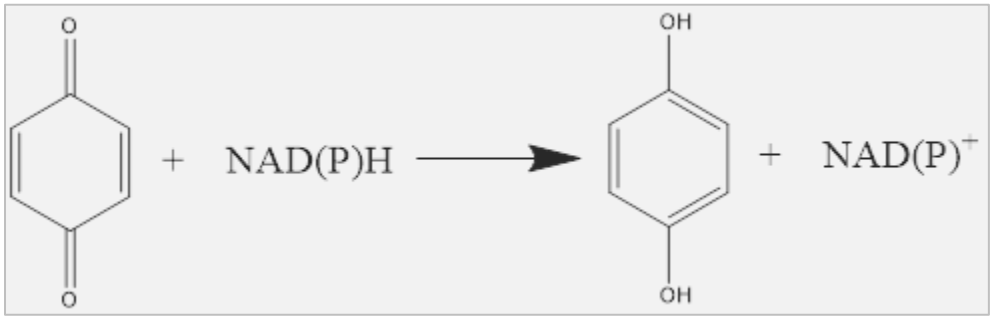
In living systems, functions are carried out by proteins and deregulation of these proteins associated with various functional pathways, are responsible for many diseases including cancer. Worldwide, cancer is responsible for almost ten million deaths annually and the burden is increasing due to poor clinical outcome, drug resistance, tumor heterogeneity, late diagnosis, disease relapse and failed clinical trials [1]. The wealth of information generated by large scale genomic data and lowering costs of Next Generation Sequencing (NGS) technologies has helped the scientific community to understand the underlying genetic causes of cancer [2]. To date, genomic and transcriptomic studies have helped us to understand many differences in cancer cells caused by genetic variation including copy number changes, gene mutations, and differential gene expression associated with specific disease subtypes that contribute to disease pathology [3]. But there has been limited success in translating this information from the bench to the clinic to develop efficient therapeutics [4]. For a systems level disease like cancer, understanding the mechanism of disease progression and cellular signaling changes is important and needs to be studied at the protein and signaling levels. Moreover, the drugs approved for cancer therapy target at the protein level and not at the gene level [5]. The regulation of protein function, degradation, localization and post-translational modifications (PTMs) play a critical role in disease progression and diagnosis [6]. PTMs are covalent modifications that change the properties of an amino acid or protein. They reversibly or irreversibly alter the structure and functional properties of a protein through biochemical reactions [7]. Various types of PTMs like phosphorylation, glycosylation, acetylation, methylation and ubiquitination are identified [8]. These changes happen downstream of gene expression, and transcript

expression measured from NGS studies do not correlate with the protein abundance present in cells. Phosphorylation is a post-translational modification that relays external stimuli through a cascade of signaling events to elicit appropriate biological responses [7]. With the emergence of Mass Spectrometry (MS) based technology, we are able to quantify protein abundance and identify PTMs like phosphorylation by identifying characteristic mass differences between modified and unmodified peptide [9]. Phosphoproteome studies have become an important tool in precision medicine to characterize the role of kinases and phosphatases in disease progression, and are targeted for disease therapy by small molecule inhibitors of protein kinases and related molecules [10].

Phosphorylation is mediated by protein kinases which transfer the gamma phosphate group from ATP to specific amino acids, mainly Serine (Ser), Threonine (Thr) or Tyrosine (Tyr) residues [11]. Every phosphorylation event adds a mass of 80 Daltons to the total mass of the peptide. The changes in phosphorylation can be studied by mass spectrometry by measuring the mass difference of the phosphorylated residues on the peptides to determine the site of phosphorylation on the protein and understand the functional consequences of that modification [12]. This new layer of information must be integrated with other *omics* studies to capture the functional state and dynamic properties of a cell.

In this study, MS based phosphoproteome analysis is performed on Triple Negative Breast Cancer (TNBC) cells upon treatment with NAD(P)H Quinone Dehydrogenase 1 (NQO1) bioactivable drug, isobutyl-deoxynyboquinone (IB-DNQ) and Poly (ADP-ribose) polymerase (PARP) inhibitor Rucaparib. NQO1 gene encodes NAD(P)H dehydrogenase enzyme in humans [13]. It is a FAD binding protein that functions as a homodimer and

catalyzes two electron reduction of toxic quinones to hydroquinones (HQ) [14]. The HQ binds to glutathione forming intracellular HQ-GSH conjugate that can be excreted from the cells [12] (Picture 1). The enzyme works with equal efficiency with both cofactors NADH and NADPH and requires a tightly bound FAD cofactor [15]. Certain compounds such as antitumor quinones can be bioactivated via the two electron reduction by NQO1 [16]. Because of the high levels of NQO1 in certain tumors, this has led to an interest in designing compounds that can be bioactivated by NQO1 as antitumor agents [17].



Picture 1: The reaction catalyzed by NQO1 in which a quinone is reduced to a hydroquinone by the NADH or NADPH

IB-DNQ is a deoxyniboquinone compound that is dependent on NQO1 for its activation and therefore is bioactivated by NQO1. NQO1 performs a two-electron reduction of the parent IB-DNQ to its HQ form, which is unstable, the HQ spontaneously reverts back to its parental, oxidized form through two reactions with molecular oxygen, the intermediate being a semiquinone (SQ) [18]. This futile cycle generates two superoxide moieties which eventually generate massive H_2O_2 concentration, which carry out DNA damage and ER Ca^{2+} release [19] (*Figure 1*).

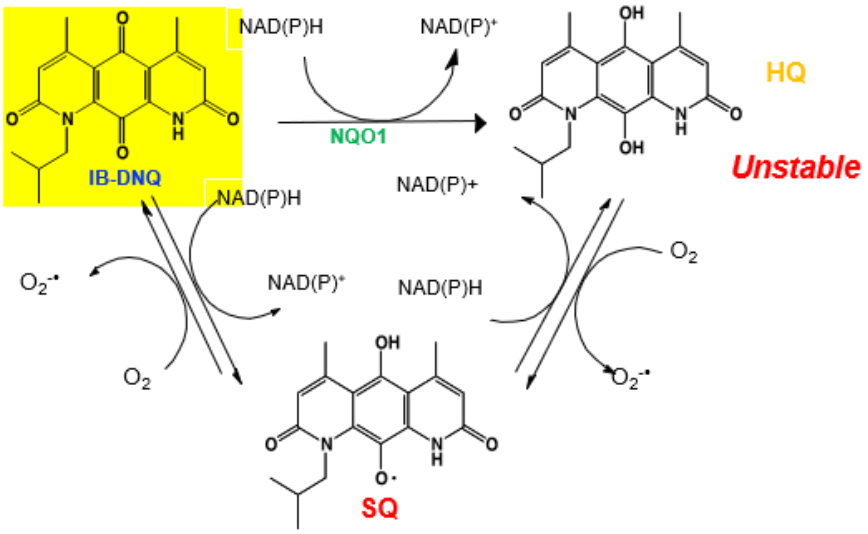


Figure 1: NQO1 mediated bioactivation of IB-DNQ

NQO1 is overexpressed in various types of solid cancers including breast cancer while catalase is overexpressed in normal cells [20]. In normal cells, the H_2O_2 produced by NQO1 mediated IB-DNQ activation is removed by catalase and they are protected. While in cancer cells, the absence of catalase causes accumulation of H_2O_2 leading to DNA damage, activation of cell cycle checkpoint machinery, PARP1 hyperactivation that causes depletion of NAD+/ATP, mitochondrial dysfunction, and apoptosis inducing factor (AIF) release resulting in tumor selective cell death [21]. The absence or low abundance of catalase is a property of cancer cells is used to selectively target them with NQO1 bioactivatable compounds using a precision medicine approach [20]. Patients with reduced expression of NQO1 or with single nucleotide polymorphism (SNP) in NQO1 show resistance to chemotherapeutic drugs and particularly NQO1 bioactivatable compounds. There are two known SNP's in NQO1, namely NQO1 type 2 (NQO1*2) and NQO1 type 3 (NQO1*3). The NQO1*2 polymorphism containing C609T change leads to proline to serine substitution at position 187 of NQO1.[22] This mutant protein is catalytically

inactive due to its reduced ability to bind the FAD cofactor [23] and its catalytic activity is only 4% compared to wild type [24]. The NQO1*3 polymorphism with C465T change leads to arginine to tryptophan substitution in the protein; this substitution results in a change in rate of quinone substrate metabolism. The type 2 polymorphism is most common [25]. The NQO1 status of cancer cells is important for the IB-DNQ treatment to work. Patients with NQO1 polymorphism will not show the desired effects with IB-DNQ treatment [26].

Previous studies with another bioactivable drug beta-lapachone showed synergistic action with PARP inhibitors. [27] BRCA and PARP are housekeeping genes involved in DNA damage repair via the homologous recombination (HR) pathway. Healthy cells having two normal copies of BRCA1 and BRCA2, will have no effect from PARP inhibitor as the cells still have BRCA working in them [28]. BRCA mutated cells with single-stranded breaks can still use PARP to fix the DNA break [29]. Whereas in cancer cells with the loss of both copies of BRCA, there will be an impaired HR pathway, which eventually will lead to cell death by PARP inhibitors [30]. Likewise, the double-stranded break (DSB) induced by very high doses IB-DNQ will be difficult to repair by HR deficient cancer cells. This synergistic action of IB-DNQ mediated DSB and cell death by PARP inhibitor Rucaparib is exploited in TNBC with minimum off-target effects.

Proteomics datasets are characterized by the wide heterogeneity commonly found in biological samples making it difficult to identify important biomarkers [9]. The first step in *omics* data analysis approach is to identify affected biological processes and functions from differentially expressed proteins between disease and normal states using various computational tools like Gene Ontology (GO) and Gene Set Enrichment Analysis (GSEA).

The MS based phosphoproteomic study only provides information about the phosphorylated residues in the protein but no functional information about the kinases affecting those phosphorylation events. Here, we used a kinase prediction tool called Kinase Substrate Enrichment Analysis (KSEA) to predict the kinases involved in the phosphorylation events and to understand the mechanism of drug action in terms of changes in cell signaling. This information can be mapped onto protein-protein interaction (PPI) network and kinase-substrate interaction networks to provide a better picture of the effect of perturbation in a given biological network (*Figure 2*).

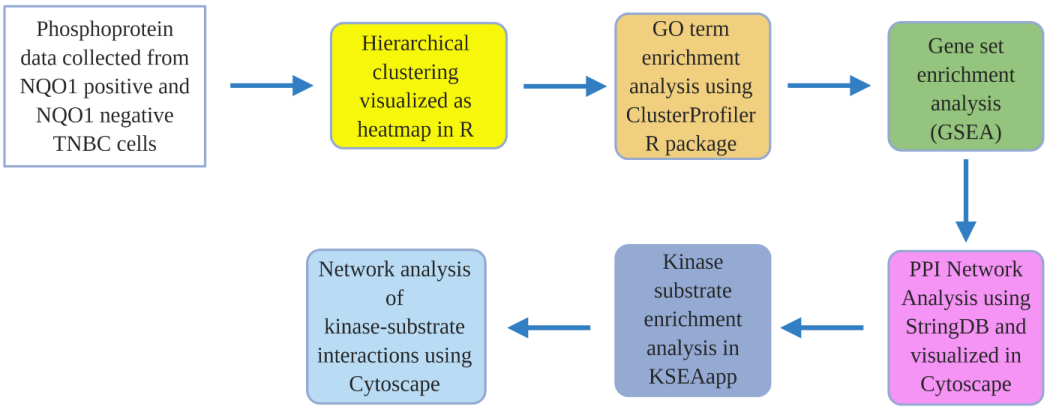


Figure 2: Flowchart of methodology used

METHODS

Proteomics of Patient-Derived TNBC Cells

This phosphoproteomic study utilized MDA-MB-231 cells which are NQO1*2 polymorphic human metastatic TNBC cells, in replicates of three. The NQO1 positive cells were created by overexpressing NQO1 by plasmid transfection. It is likely that overexpressed NQO1 maybe substantial amount higher than endogenous expression however overexpression is consistent with previously shown work which indicates overexpression of NQO1 in many solid cancers [31]. Both NQO1 positive and NQO1 negative cells were treated for 4 hours with IB-DNQ 0.1 μM (sublethal dose), IB-DNQ 0.4 μM (lethal dose), Rucaparib 15 μM and a combination therapy of IB-DNQ 0.1 μM and Rucaparib 15 μM. A previous dose response study by Huang et. al. found synergistic lethality between 15 μM Rucaparib and another NQO1 bio activable drug β-lapachone [20]. Rucaparib is chosen in this study since clinical grade formulation was available. DMSO was used as the negative control. Approximately 10 million cells were used to extract the proteins by denaturing with 8M urea, followed by LysC/Trypsin Gold (Promega) digestion. The endoprotease LysC is more stable than trypsin and is used before trypsin digestion under harsh, solubilizing conditions such as 8M urea. 45 μg of sample was used for global analysis and 3 milligrams for phosphopeptide enrichment using TiO_2 chromatography. The peptides from different treatment conditions and replicates are multiplexed and labelled with different Tandem Mass Tag (TMT) labels from Thermo Fisher. The TMT labels are isobaric having exactly the same mass but with different isotopes of Nitrogen (^{15}N) and Carbon (^{13}C). The high resolution Thermo-Orbitrap Fusion Lumos mass spectrometer can differentiate between the very small mass difference of the

isotopes of N and C. Each tag has a cleavable mass reporter which is unique and reacts with primary amines (peptide N terminus and Lysine residues). After labelling was quenched, the samples were mixed together and subjected to high pH basic fractionation in Thermo-Orbitrap Fusion Lumos instrument using MS3 acquisition. In MS2 stage, Collision Induced Dissociation (CID) is performed using helium to generate b and y fragment ions from breakage of peptide bonds. The b ions extend from the N-terminus and y ions extend from the C-terminus of the peptides. A population of those ions are selected by the Orbitrap for higher-energy collisional dissociation (HCD) fragmentation in MS3 step. This breaks off the TMT reporter ion which is picked up by the detector (*Figure 3*).

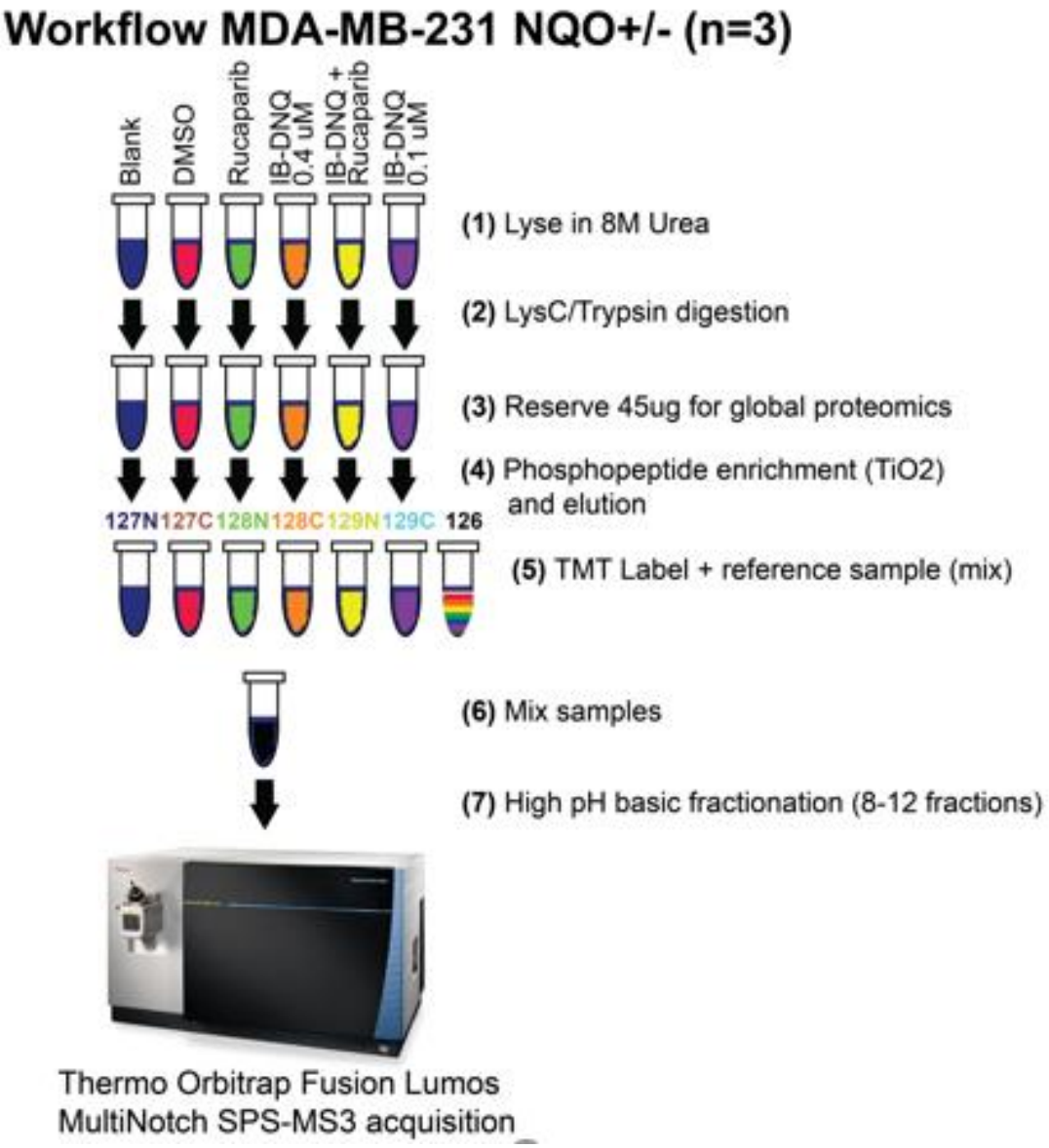


Figure 3: Workflow of proteomic analysis of MDA-MB-231 cells

Analysis of Raw Data

The raw data was analyzed in Proteome Discoverer [32] for both global and phosphoproteomic datasets. 4383 unique proteins and 22872 phosphorylation sites were quantified in NQO1 positive cells and 3181 unique proteins and 8530 phosphorylation sites were quantified in NQO1 negative cells (*Figure 4*).

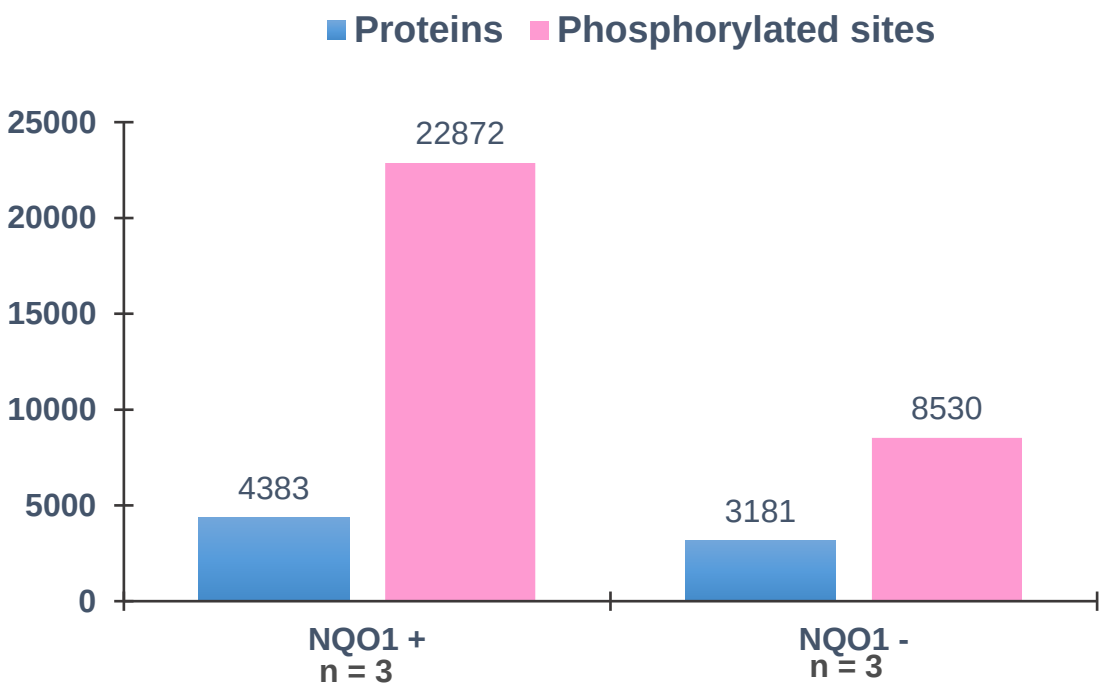


Figure 4: Number of identified proteins (blue) and phosphorylated sites (pink) in MDA-MB-231 cells

Gene Ontology (GO) Analysis

GO summarizes the long lists of proteins into biologically meaningful information. A T-test was performed on proteins differentially expressed between two experimental groups. The p-value adjustment was calculated to avoid false positives using either Bonferroni correction for small scale experiments or Benjamini Hochberg procedure for large scale experiments. The upregulated proteins with fold change ratio > 1 and down regulated proteins with fold change ratio < 1 are analyzed separately. The peptide level output from Proteome Discoverer is exported in excel format and processed in R. The UniProt protein accession numbers are converted to gene symbols using the *mapIds* function from *Annotationdbi* Bioconductor package [33] (Picture 2).

```
library('org.Hs.eg.db')
protein_id<- rawdata$protein_id
id_gene<-mapIds(org.Hs.eg.db, protein_id, 'SYMBOL',
'UNIPROT', multiVals="first")
id_entrez<-mapIds(org.Hs.eg.db, protein_id, 'ENTREZID',
'UNIPROT', multiVals="first")
id_ensembl <- mapIds(org.Hs.eg.db, protein_id, 'ENSEMBL',
'UNIPROT', multiVals = "first")
```

Picture 2: R Code

The rawdata was filtered by p value < 0.05 (Picture 3).

```
Rucaparib <- data[, c(1,23,24,25,2:5,11,17)]
Rucaparib_pvalue_sorted <- Rucaparib %>% filter(ratio.p_
Rucaparib < 0.05)
```

Picture 3: R code

The enrichment analysis was performed using the *ClusterProfiler* package in R [34]. *ClusterProfiler* is an enrichment tool for statistical analysis and visualization for functional and comparative study. It supports three species, including *H. sapiens*, *M. musculus* and *S. cerevisiae*. Large scale protein expression analysis from mass spectrometry-based experiments yields many significant proteins. To understand their biological relationships Gene Ontology (GO) enrichment analysis was performed with the *enrichGO* function [35]

from *ClusterProfiler* (Picture 4). The enriched GO terms are visualized in the form of barplot using the R base function *barplot*. The bars are colored based on the adjusted p-value calculated.

```
Rucaparib_up1 <- Rucaparib %>% filter(ratio.p_Rucaparib <
0.05 & abs(ratio_Rucaparib) > 1.0)
up_Ruca_BP <- enrichGO(Rucaparib_up1 [,7], OrgDb =
org.Hs.eg.db,
keyType = "SYMBOL",ont="BP",  pAdjustMethod = "BH",
pvalueCutoff=1, qvalueCutoff=1)
Rucaparib_down1 <- Rucaparib %>% filter(ratio.p_Rucaparib
< 0.05 & abs(ratio_Rucaparib) < 1.0)
down_Ruca_BP <- enrichGO(Rucaparib_down1 [,7], OrgDb =
org.Hs.eg.db, keyType = "SYMBOL",ont="BP",
pvalueCutoff=1, qvalueCutoff=1)
```

Picture 4: R Code

Gene Concept Network

While the barplots displays only the enriched terms, to view the genes associated with those terms the function *cnetplot* [36] from the package enrichplot is used [37]. Some genes may be associated with more than one biological pathway, which is difficult to visualize on a barplot. The network plot shows such linkages (Picture 5).

```
enrichplot::cnetplot(up_Ruca2) + ggtitle("NQO1+ Network
plot of enriched terms-Rucaparib") + theme(plot.title =
element_text(hjust = 0.5, size=20,face="bold"))
```

Picture 5: R Code

Gene Set Enrichment Analysis (GSEA)

The GSEA doesn't require cutoffs and therefore is threshold free and is more powerful than GO. The use of cutoffs introduces bias in the hypergeometric tests. The whole list of genes/proteins detected in the phosphoproteomics experiment was used. The goal was to determine whether the members of the gene/protein set were randomly distributed throughout a ranked protein list or if they were located at the top or bottom of the ranked list. The list of proteins was sorted based on T statistic like p value < 0.05 and the enrichment score (ES) is calculated using the T statistic. The ES for a set of proteins is the maximum value reached and that may be positive or negative. Thereafter permutations of the samples were performed to recalculate random ES scores. The ES was normalized (NES), and adjusted p value was estimated based on randomized ES (Picture 6).

$$\text{NES} = \frac{\text{actual ES}}{\text{mean (ES against all permutations of the dataset)}}$$

Picture 6: Normalized enrichment scores

Here the GSEA Java desktop application from the Broad Institute was used to summarize the long list of proteins into biological pathways [38]. Two types of input files are required

for GSEA analysis; an expression dataset in *gct* format and phenotype annotation file in *cls* format. The *cls* file is required only for *gct* 1.2 version. The hallmark gene sets from MSigDB (Molecular Signatures Database) in *gmt* format was used for analysis [39]. The hallmark gene sets were generated by computational methods and summarize well defined biological states and processes. These gene sets reduce noise and redundancy in GSEA analysis. The first line in a *gct 1.2* file format is always #1.2. The second line contains two numbers representing the number of rows and number of columns in the data matrix (*Figure 5*). Since the GSEA analysis is gene centric, gene ID's in the form of HGNC symbols are used. Alternatively, the *gct* file can be generated using Morpheus (https://software.broadinstitute.org/morpheus/) by uploading the metadata in excel format [40]. Upload the excel file with HGNC gene symbols and log2 fold change values, change row and column annotations from View → Options → Annotations. Since gene centric pathway analysis requires single entries of genes, Morpheus can collapse the dataset to unique gene symbols from Tools → Collapse → select method to median → Collapse rows → by id. A new tab will be created with a gene centric heatmap which can be downloaded in *gct 1.2* format.

#1.2						
3234	5					
Name	Description	Rucaparib	IB_DNQ_Rucaparib	IB_DNQ_0.1M	IB_DNQ_0.4M	DMSO_control
UBR4	Q5T4S7	-0.15	-0.22	-0.43	0.15	0.14
RBM15	Q96T37	-0.03	-0.18	-0.59	-0.36	0.14
RRP9	O43818	-0.04	0.69	-0.32	0.44	0.04
NOLC1	Q14978	-0.2	-1.23	-0.64	-0.49	-0.33
NUCKS1	Q9H1E3	-0.18	-0.73	-0.42	-0.05	0.19
HSP90AA1	P07900	0.17	-2.03	-0.02	-0.69	0.31
TLE3	Q04726	0.3	0.39	-0.2	0.11	0.18

Figure 5: gct 1.2 file format

The *cls* file format contains phenotype labels of each sample in the dataset, and values are separated by tab or space. The first line consists of three numbers; the number of samples, number of classes of samples and the third number is always 1. The second line begins with a pound sign (#) and contains the class names (e.g. wild type vs. mutant; drug vs. control). The third line contains class labels present in the dataset (*Figure 6*).

```
5 5 1
# Drug1 Drug2 Drug3 Drug4 Control
Rucaparib IB_DNQ_Rucaparib IB_DNQ_0.1M IB_DNQ_0.4M DMSO_control
```

Figure 6: cls file format

The *gct* and *cls* files are loaded in GSEA using 'Load data', 'Run GSEA' option lets the user choose the expression dataset (*gct* file), Gene set database (Hallmark gene sets in *gmt* format from dropdown menu) and Phenotype labels (*cls* file). Number of permutations (1000), no collapse of gene symbols, Permutation type (gene sets), Chip platform to be kept blank. Rename the analysis and save results in desired folder keeping other options as default. The permutations mean repeating the analysis 1000 times and creating a distribution of ES scores. The significant ES scores are used to calculate p-values. The 'Run' option will run the analysis and generate GSEA reports on the left panel.

Computational Method for Kinase Enrichment Analysis

Various computational tools are available to decipher the kinases associated with the identified phosphorylation sites. But all these tools have certain drawbacks with respect to the organisms they support and the number of kinases they can predict. Majority of such

tools require complete fasta sequences of proteins rather than individual phosphorylation sites detected from MS based phosphoproteomic experiments (*Table 1*).

Tool	Application	Organism	Input
NetworKIN	pinpoints kinases responsible for specific phosphorylations	human and yeast only	complete FASTA sequence
KEA3	computes enrichment of a query protein set over kinase-substrate sets	NA	gene list
PSEA	GSEA inspired method adapted for proteomics	human, yeast, Drosophila, rat, S. cerevisiae	input file with spectrum count values or abundance ratio
GPS 5.0	eukaryotic protein kinases and protein phosphatases database	164 species	complete FASTA sequence
KinasePhos 2.0	Identifies human kinase-regulated phosphorylation sites	human	complete FASTA sequence
PhosphoPICK	predicts kinase substrates	107 human kinases, 24 mouse kinases and 26 yeast kinases	complete FASTA sequence
NetPhos	Predicts serine, threonine or tyrosine phosphorylation sites in eukaryotic proteins. Predictions are made for the following 17 kinases: ATM, CKI, CKII, CaM-II, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK, SRC, cdc2, cdk5 and p38MAPK.	euakryotes	complete FASTA sequence
Quokka	Identifies human kinase-regulated phosphorylation sites	human	complete FASTA sequence

Table 1: Computational tools available for phosphorylation site and kinase prediction

Kinase Set Enrichment Analysis or KSEA is an approach to infer the kinase activity from phosphoproteomics dataset based on individual phosphorylation sites. KSEA is published as a freely available Python code (Python version 2.7.x) [41] and also as an R shiny app [41]. Both of these resources infer the kinase activity by utilizing prior knowledge of kinase-substrate relationships from curated databases or computational prediction tools. Here the R shiny app version of the tool called KSEAapp was used, available as an R package on CRAN: CRAN.R-project.org/package¼KSEAapp/. Currently the KSEA tool only supports human kinases but other organisms can also be analyzed by extracting the species-specific kinase-substrate entries from the dataset available on PhosphoSitePlus.

KSEA Algorithm Overview

The KSEA algorithm works on the phosphoproteomics dataset with test and control samples. The algorithm takes an input file in *csv* format consisting of a list of peptides, corresponding protein and gene names, phosphosites and associated fold change (FC) and p values. The gene names, phosphosites and FC (not log transformed) values are necessary for the calculations. The protein, peptide and p values are optional, but the headers must be present in the input file with the rest of the column as 'NULL' but not left blank. KSEA uses kinase substrate relationships from both PhosphoSitePlus [42] and NetworkIn [43], and users have a choice of selecting kinase-substrate relationship from PhosphoSitePlus alone or both PhosphoSitePlus and NetworkIn to improve coverage or get more kinases. The p-value cutoff and substrate count cutoff are required for the bar plot and was set at 0.05 and 2, respectively. The value of 2 for substrate count cutoff means any kinase with a minimum of 2 substrates will be included in the barplot. The algorithm scores each kinase based on the relative hyperphosphorylation and dephosphorylation of its substrates, as identified from the phosphosite specific databases such as PhosphoSitePlus and NetworkIn. The positive and negative values represent an increase or decrease in kinase activity in response to drug treatment relative to control. The kinase is scored based on the formula:

$$score = \frac{(\bar{s} - \bar{p})\sqrt{m}}{\delta}$$

Picture 7: Kinase score formula

Where $\bar{s}$ denotes the mean $\log_2$(FC) of known phosphosite substrates of the given kinase, $\bar{p}$ represents the mean $\log_2$(FC) of all phosphosites in the dataset, m denotes the total number of phosphosite substrates identified from the experiment that annotate to the specified kinase, and δ denotes the standard deviation of the $\log_2$(FC) across all phosphosites in the dataset. This formula is based on a z-score transformation, and the resulting scores (denoted as 'z-score' in the KSEA App outputs) are assumed to be normally distributed. Subsequently, the p-value is determined by assessing the one-tailed probability of having a more extreme score than the one measured, followed by a Benjamini-Hochberg FDR correction for multiple hypothesis testing [44]. The Benjamini-Hochberg procedure is a statistical tool to control false discovery rate (FDR) in multiple hypotheses testing and control false positives.

Input File Preparation

The phosphoproteomic dataset for NQO1 positive cells was prepared using R. The columns 'Master protein accessions', 'Sequence', 'Modifications in master proteins', 'Abundance ratio' and 'Abundance ratio p-value' for different treatment conditions were chosen for KSEA analysis. The column 'Modifications in master proteins' was edited in excel itself to remove all texts except the modified phosphosites, multiple phosphosites were separated by a semicolon without any spaces. Since KSEA only takes HGNC gene symbols, the UniProt identifications from the phosphoproteome dataset were converted to HGNC gene symbols using the function *mapIds* from Annotationdbi package in Bioconductor (Picture 8).

```
library('org.Hs.eg.db')
protein_id <- rawdata$protein_id
id_gene<-mapIds(org.Hs.eg.db, protein_id, 'SYMBOL',
'UNIPROT', multiVals="first")
Rucaparib <- data[, c(1,21,2,4,16,6)]
Rucaparib_pvalue_sorted <- Rucaparib %>%
filter(ratio.p_Rucaparib < 0.05 &
!is.na(ratio.p_Rucaparib) & !is.na(id_gene) &
!is.na(`Modifications in Master Proteins`))
```

Picture 8: R Code

Finally, the columns are named as per the requirements of the KSEAapp algorithm (Picture 9).

```
colnames(Rucaparib_pvalue_sorted)<-c("Protein", "Gene",
"Peptide","Residue.Both", "p", "FC")
```

Picture 9: R Code

Sequence Motif Analysis

The specificity of protein kinases is associated with the consensus sequence surrounding the phosphorylated serine, threonine and tyrosine residues. The motif analysis is performed with the *ggseqlogo* R Bioconductor package [45]. *ggseqlogo* takes two types of input: a character vector of sequences or position frequency matrix (PFM) or Position-Specific

Scoring Matrix (PSSM), where the rows are amino acids and the columns are the positions of those amino acids. The peptide sequences from kinase-substrate relationships from the KSEA analysis was used as input (Picture 10).

```
ggseqlogo(PLK1_data, seq_type='aa')
```

Picture 10: R Code

The *ggseqlogo* algorithm builds a matrix of all the peptide sequences with amino acids on the rows and the positions correspond to the alignment columns. The algorithm calculates the probability of observing an amino acid at that position. The dash means the amino acid is not present in the peptide sequence (*Table 2, Table 3*).

Position	1	2	3	4	5	6	7	8	9	10	11
Sequence 1	G	D	V	E	G	S	Q	S	Q	D	E
Sequence 2	D	E	G	E	G	S	G	E	S	E	R
Sequence 3	S	P	S	L	D	S	G	D	S	G	S
Sequence 4	L	D	S	G	D	S	D	S	E	E	L
Sequence 5	E	E	D	S	K	S	I	K	S	D	V
Sequence 6	R	M	E	E	E	S	Q	S	Q	G	R
Sequence 7	E	E	E	S	Q	S	Q	G	R	D	L
Sequence 8	G	M	D	L	N	S	D	D	S	T	D
Sequence 9	L	N	S	D	D	S	T	D	D	E	A
Sequence 10	N	S	D	D	S	T	D	D	E	A	H
Sequence 11	S	R	K	R	R	S	S	T	V	A	P
Sequence 12	P	G	K	Q	D	T	E	E	D	E	E
Sequence 13	E	P	G	N	G	S	L	D	L	G	G

Table 2: Matrix of PLK1 peptide sequences; the rows represent peptide sequences and columns are positions of amino acids in the sequence

Position	1	2	3	4	5	6	7	8	9	10	11
Prob (G)	0.15	0.08	0.15	0.08	0.23	--	0.15	0.08	--	0.23	0.08
Prob (A)	--	--	--	--	--	--	--	--	--	0.15	0.08
Prob (V)	--	--	0.08	--	--	--	--	--	0.08	--	0.08
Prob (L)	0.15	--	--	0.15	--	--	0.08	--	0.08	--	0.15
Prob (I)	--	--	--	--	--	--	0.08	--	--	--	--
Prob (M)	--	0.15	--	--	--	--	--	--	--	--	--
Prob (W)	--	--	--	--	--	--	--	--	--	--	--
Prob (F)	--	--	--	--	--	--	--	--	--	--	--
Prob (P)	0.08	0.15	--	--	--	--	--	--	--	--	0.08
Prob (S)	0.15	0.08	0.23	0.15	0.08	0.85	0.08	0.23	0.31	--	0.08
Prob (T)	--	--	--	--	--	0.15	0.08	0.08	--	0.08	--
Prob (C)	--	--	--	--	--	--	--	--	--	--	--
Prob (Y)	--	--	--	--	--	--	--	--	--	--	--
Prob (N)	0.08	0.08	--	0.08	0.08	--	--	--	--	--	--
Prob (Q)	--	--	--	0.08	0.08	--	0.23	--	0.15	--	--
Prob (D)	0.08	0.15	0.23	0.15	0.31	--	0.23	0.38	0.15	0.23	0.08
Prob (E)	0.23	0.23	0.15	0.23	0.08	--	0.08	0.15	0.15	0.31	0.15
Prob (K)	--	--	0.15	--	0.08	--	--	0.08	--	--	--
Prob (R)	0.08	0.08	--	0.08	0.08	--	--	--	0.08	--	0.15
Prob (H)	--	--	--	--	--	--	--	--	--	--	0.08

Table 3: Position-Specific Scoring Matrix (PSSM) of PLK1 peptide sequences

A sequence logo is a graphical representation of the aligned sequences, where the size of each residue is proportional to its frequency at that position and total height of all the residues in the position is proportional to the conservation of the position. The height of the amino acids is measured by the entropy and is depicted as bit score shown on the Y axis and is calculated using the formula: (Picture 11)

$$H(l) = -\sum_{b=Ala}^{Val} f(b,l)\log_2 f(b,l)$$

$$R_{sequence}(l) = \log_2 20 - \left(H(l) + e(n)\right)$$

Picture 11: Sequence motif formula

The frequency of an amino acid at a position is calculated and multiplied by $\log_2$ of the frequency of that amino acid at that position. This value is subtracted from the $\log_2 20$ in the case of protein sequences, as there are 20 amino acids. H is Shannon entropy. R is 4.32 bits in case of protein sequences, which is $\log_2(20)$. This strategy can be used to discover motifs from sets of phosphorylation sites detected in phosphoproteomic experiments.

Visualization of Kinome Data

To visualize the kinome data identified by Kinase Enrichment Analysis, an interactive R shiny web application called CORAL was used [46]. CORAL allows the visualization of both qualitative and quantitative data in three modes: the traditional kinome tree, based on Cell Signaling Technology, the radial network with the nodes representing kinase, groups, families and subfamilies, and dynamic force networks. The node color, node size and branch color can be customized according to user requirements. The quantitative kinase enrichment data with corresponding log2-fold change values is used as input to highlight differentially expressed kinases or highlight targets of kinase inhibitors. The Shiny web application is available at http://phanstiel-lab.med.unc.edu/CORAL/. In the *Plot* tab, color scheme is selected as *Quantitative* for all the parameters and the kinase HGNC symbols and corresponding $\log_2$-fold change values from KSEA analysis are entered in separate rows. The plots are automatically populated based on the input data and nodes colored according to the fold change difference between the kinases.

Protein-Protein Interaction Analysis

The cellular signaling process is dependent on interaction between the proteins which are highly specific physical interactions. These interactions are controlled by electromagnetic and electrostatic forces, hydrogen bonding and hydrophobic effect. Understanding PPIs is important for comprehending cellular process and disease microenvironment. StringDB was used for protein interaction analysis and the corresponding fold change values are overlaid on the network with Cytoscape for visual representation of the up and downregulated proteins under different treatment conditions. The UniProt accessions were imported under the 'STRING: protein query' as data source to create the PPI network. The nodes are colored according to the fold change values by importing the excel data table directly. To understand the systems level interaction of the identified proteins with other proteins in the network, additional interactions from PPI databases can be imported while creating the String network.

Kinase Substrate Interaction Network

The kinase and phosphosite correlation obtained from the KSEA analysis was used to infer the kinase and substrate regulatory relationship and construct a kinase-substrate interaction network in breast cancer. The kinase-substrate link output file from KSEA can be directly used as input to create the network in Cytoscape [47]. Cytoscape is a versatile program to create a customized network from any type of interaction data and provides a range of applications as plugins for layering. The kinase gene was set as the source node and the substrate gene set as target node, substrate modification can be set as target node attribute for layering to create more informative networks, $\log_2$FC was selected as the edge attribute.

The resulting network was converted into a String network from Apps → String → Stringify network with a confidence score cutoff of 0.70. Note that UniProt identifications provide better results with String database [48] and therefore, kinase gene names must be converted to UniProt accession numbers. The kinase information from public database is layered by importing kinase-substrate network from PubMed, File → Import Network from Public Databases → Data Source (String: PubMed Query) → Species (Homo sapiens) → keyword (kinase) → confidence score cutoff of 1.00 → Maximum number of proteins = 100. The networks are merged to create an intersection network. Cytoscape also allows interaction data from public primary PPI databases like IntAct, MINT, BioGrid, IMEx part from String.

RESULTS AND DISCUSSION

Global and Phosphoproteome Analysis of NQO1 Positive MDA-MB-231 Cells

NQO1 was detected in the NQO1 positive samples but it was not detected in the NQO1 negative samples, however there was no statistically significant change in NQO1 abundance in NQO1 positive samples. The hierarchical clustering of proteins differentially expressed relative to DMSO in four samples of MDA-MB-231 cells treated with 0.4 µM IB-DNQ, 0.1 µM IB-DNQ, 15 µM Rucaparib and a combination therapy of 0.1 µM IB-DNQ and 15 µM Rucaparib shows separation of protein expression in different clusters. Protein intensity is presented in logarithmic scale with lower intensity color coded in red and high intensity colored in green. The rows represent individual proteins and the columns represent samples. A large number of proteins are significantly changed in combination therapy with both upregulation and downregulation and a few changes in 0.4 µM IB-DNQ compared to Rucaparib 15 µM and low dose 0.1 µM IB-DNQ. It is clearly visible from the heatmap that proteins that are upregulated in combination therapy (green) show different expression patterns in IB-DNQ 0.4 µM (colored red) indicating that the combination therapy causes more perturbations in protein expression than IB-DNQ alone. Also, very few changes are observed in single agent Rucaparib treatment compared to all other drug concentrations (*Figure 7*).

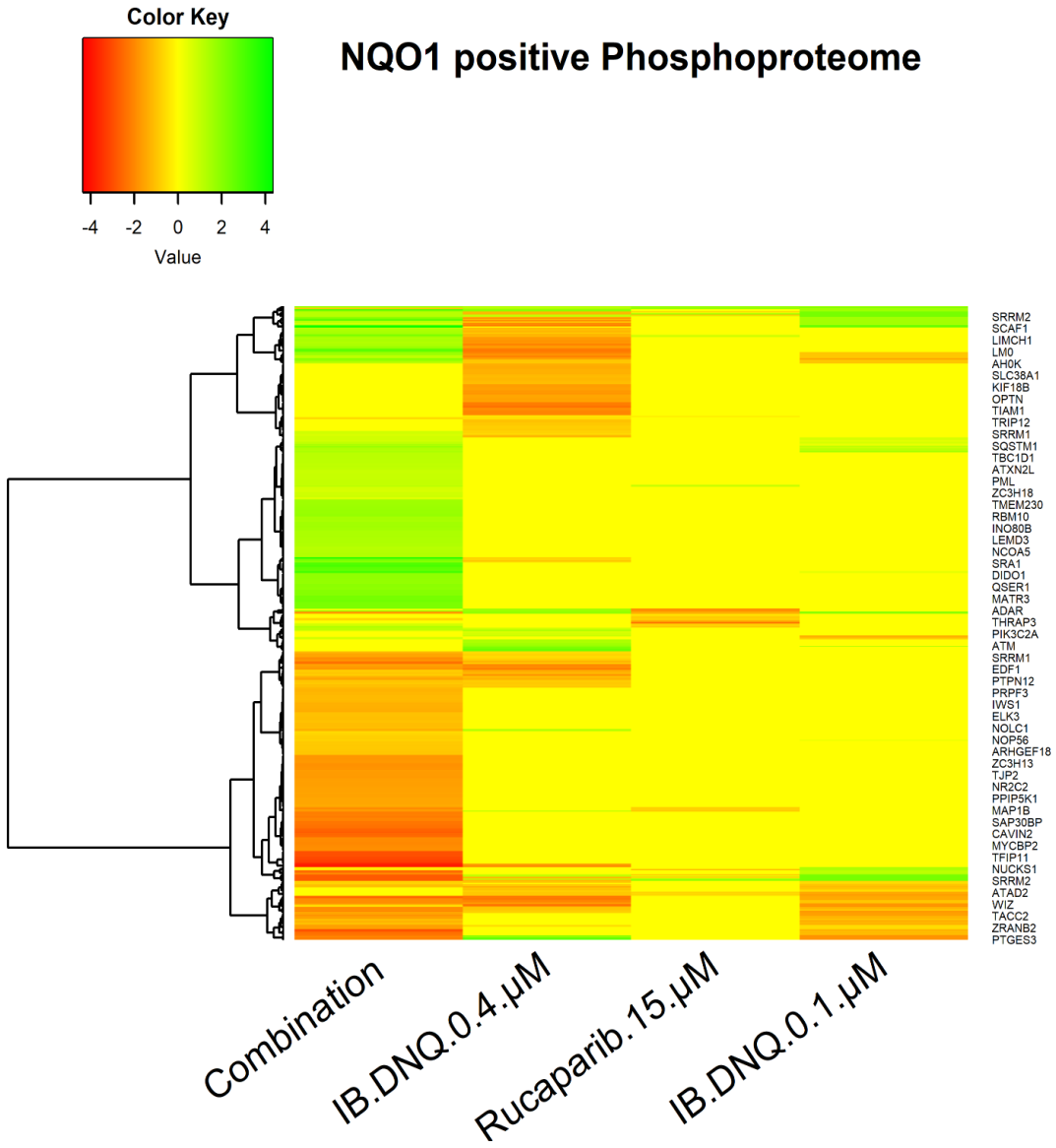


Figure 7: Hierarchical clustering of differentially expressed protein from phosphoproteome experiment in NQO1 positive cells

The volcano plots allow for quick identification of genes with large fold changes that are also statistically significant in terms of p value. Each dot represents a protein and the outliers represent most highly differentially expressed proteins. TNBC cells reveals clear differences in global protein levels between NQO1 positive (+) and NQO1 negative (-)

cells. Very few changes are observed in NQO1 negative cells (as compared to the DMSO control), while in NQO1 positive cells, only one protein is found to be significant in single agent Rucaparib 15 µM treated cells (marked in red). Significant changes are observed in cells treated with the combination therapy (0.1 µM IB-DNQ and 15 µM Rucaparib) and 0.4 µM IB-DNQ treatment (*Figure 8*). As IB-DNQ is bioactivated by NQO1, the status of NQO1 is responsible for the changes observed in NQO1 positive and NQO1 negative cells. Since larger changes are observed in NQO1 positive cells in response to IB-DNQ alone and in combination with Rucaparib, our analysis here is focused on the changes relative to DMSO in NQO1 positive cells only. In the case of peptide level changes in the phosphoproteome analysis with different treatments of IB-DNQ and Rucaparib, few changes are observed in single agent 15 µM Rucaparib. In the case of 0.4 µM IB-DNQ and combination therapy drugs, significant changes are observed with more pronounced effects in the latter. Downregulation of POLR2A C- terminal domain (CTD) and IWS1, a histone chaperone protein involved in transcription elongation indicate suppression of transcription machinery and upregulation of H2AX and XPC phosphorylation indicates DNA damage due to generation of free radicals by NQO1 mediated IB-DNQ bioactivation (*Figure 9*).

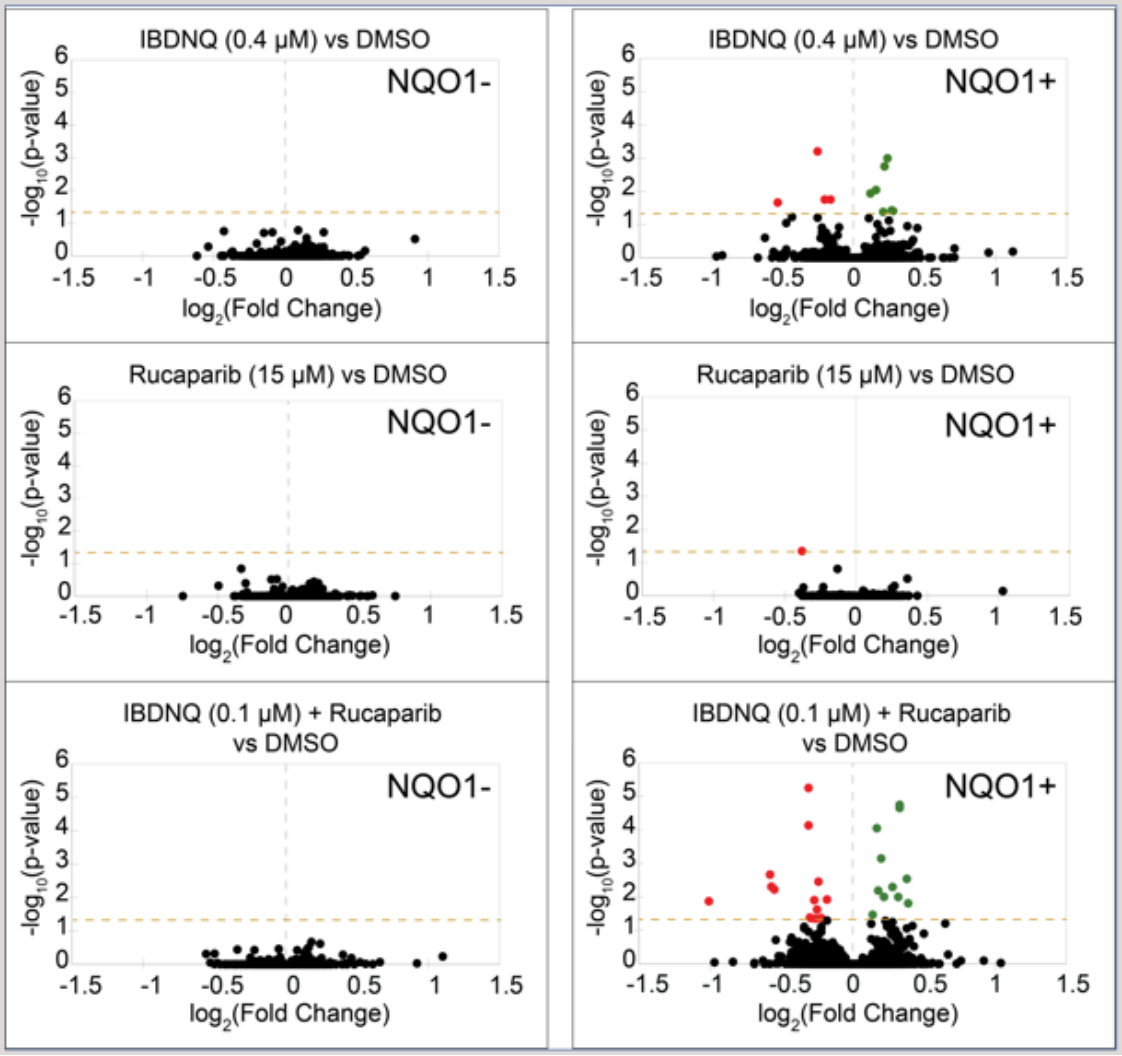


Figure 8: Volcano Plots showing comparison of global protein levels in NQO1 positive and NQO1 negative cells

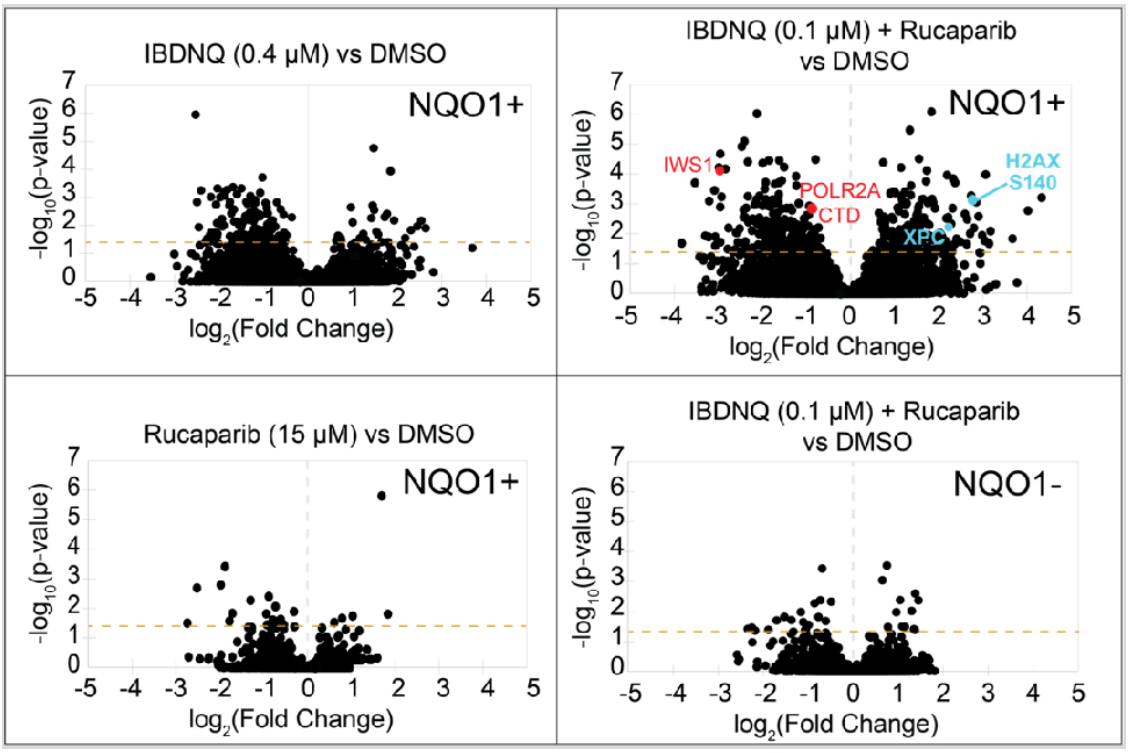


Figure 9: Volcano Plots showing comparison of phosphopeptide level changes in NQO1 positive and NQO1 negative cells

Gene Ontology Analysis of Phosphoproteome Dataset

GO term analysis reveals enrichment for mRNA processing, ATP dependent chromatin remodeling, RNA polymerase II binding and SWI-SNF complexes among others (*Figures 10-12*).

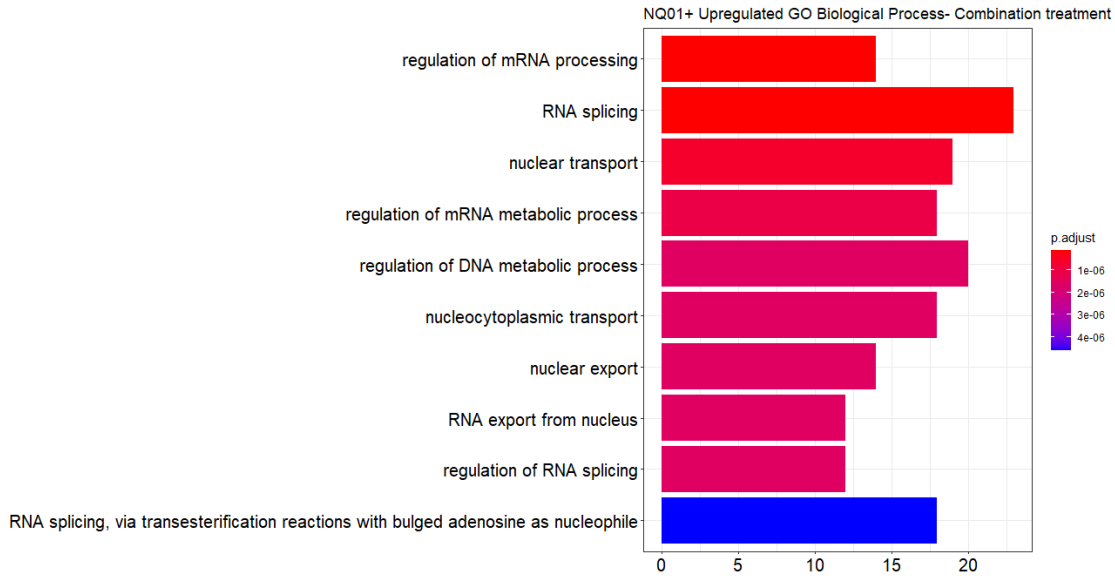


Figure 10: Upregulated Gene Ontology (GO) terms for Biological Process in NQO1 positive cells treated with combination drug therapy

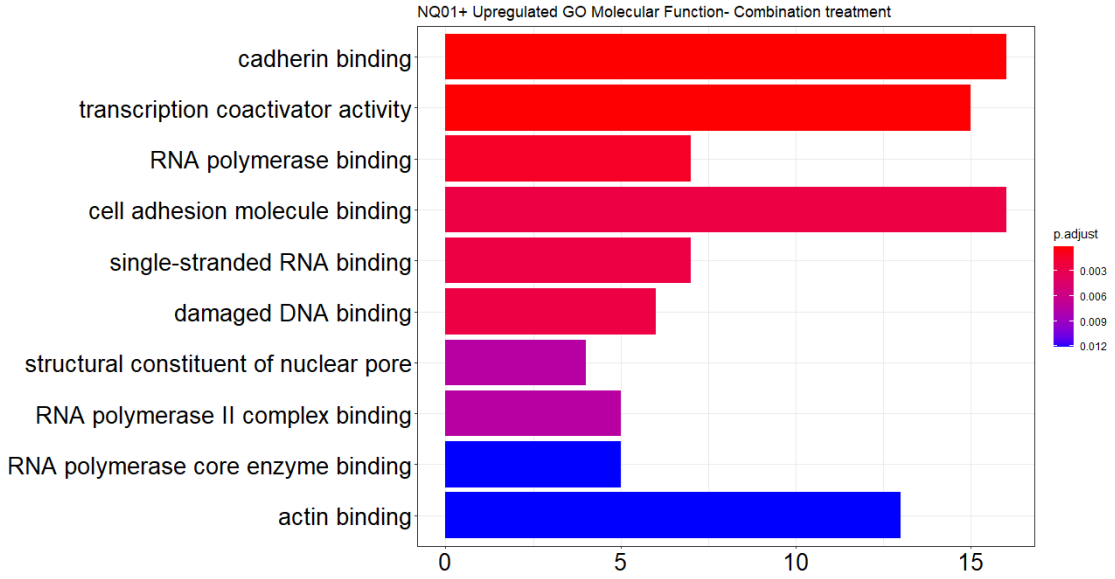


Figure 11: Upregulated Gene Ontology (GO) terms for Molecular Function in NQO1 positive cells treated with combination drug therapy

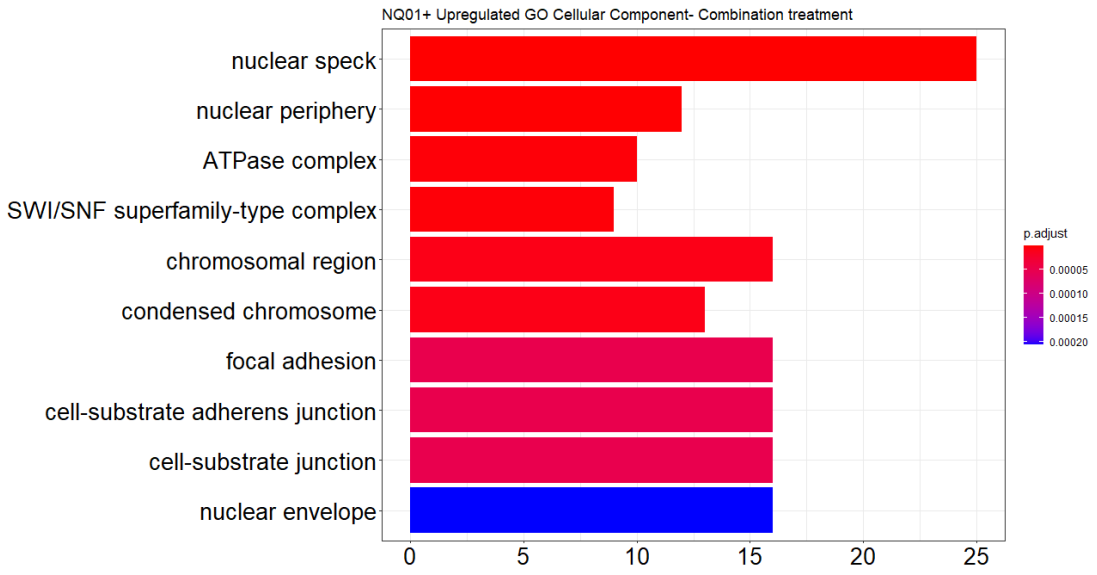


Figure 12: Upregulated Gene Ontology (GO) terms for Cellular Component in NQO1 positive cells treated with combination drug therapy

The barplots shown above only display the significantly enriched terms. The gene concept network shows the genes or proteins that are associated with those terms. Sometimes genes may be associated with multiple biological pathway categories and to view this complex association, a network plot is useful. For combination drug treated cells, we see downregulation of POLR2A, IWS1, XRCC1, TP53BP1, and KAT7 indicating suppression of DNA damage response proteins. XRCC1 and TP53BP1 are involved in DNA double stranded break repair pathways while KAT7 (histone acetyltransferase KAT7) is responsible for ATR phosphorylation and histone acetylation to recruit XPC at the DNA damaged sites in nucleotide excision repair (NER) pathway. IWS1 is a histone chaperone protein involved in transcription elongation, and POLR2A indicates downregulation of transcription machinery in response to drug treatment with a combination of IB-DNQ 0.1 µM and Rucaparib 15 µM *(Figure 13).* Also, upregulation of proteins involved in DNA

damage, unfolded protein response (UPR) and inhibition of transcription machinery agrees with the hypotheses that the combination of NQO1 bio-activable drug and PARP inhibitor is more effective in killing cancer cells. For example, H2AX is a variant of histone H2A which plays a central role in transcriptional regulation by checkpoint mediated cell cycle arrest in response to DSBs. It is phosphorylated at Ser-140 to form gamma-H2AX or H2AX-139ph and is an indication of DNA damage [49]. HNRNPU inhibits transcription elongation by inhibiting the C-terminal domain (CTD) of POLR2A, and PSMD4 is responsible for maintaining protein homeostasis by removing misfolded or damaged proteins. PML and SP100 plays a role in positive regulation of p53/TP53 and double stranded break repair and is found to be enriched in combination therapy *(Figure 14).*

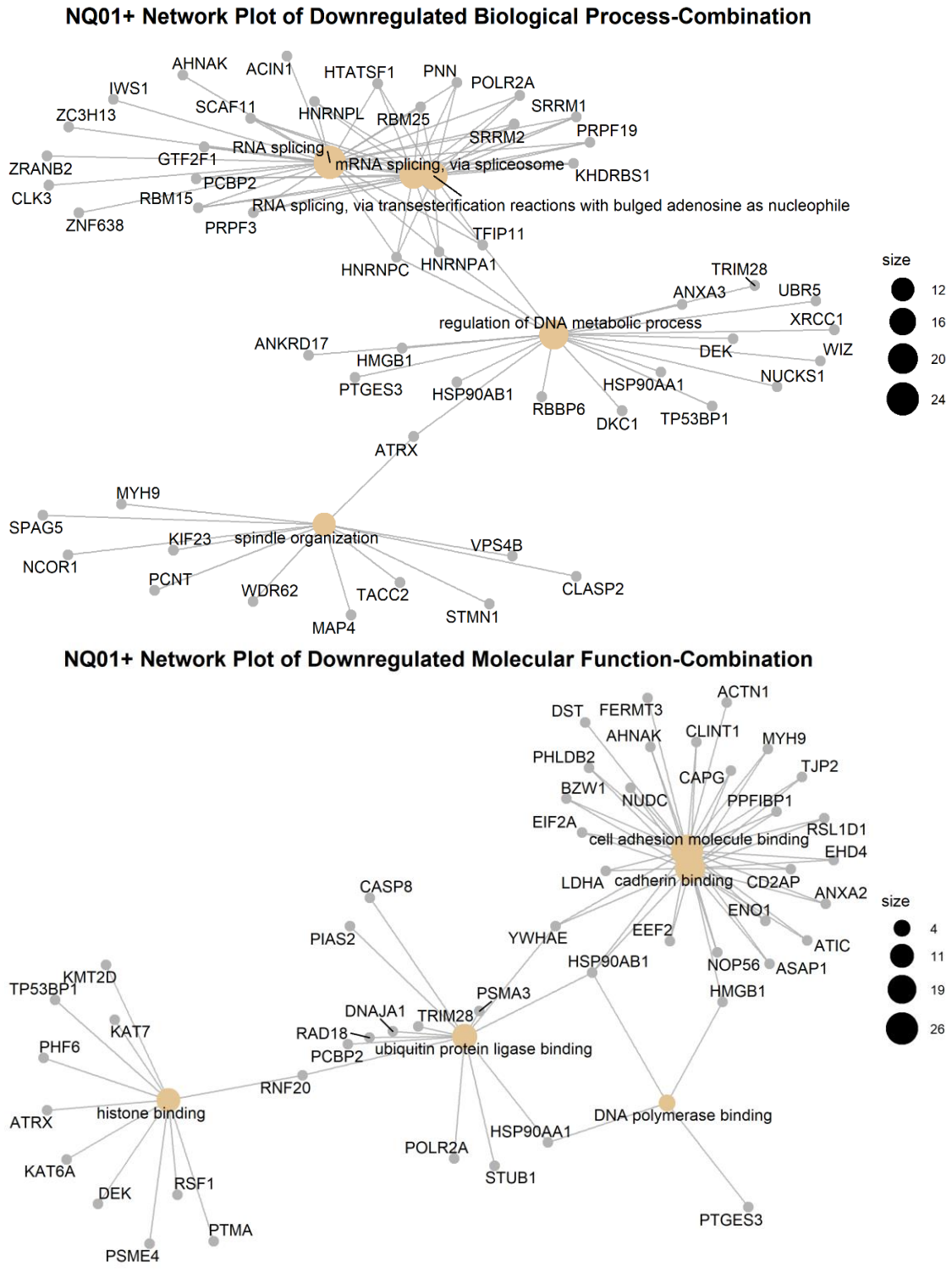


Figure 13: Gene Concept Network of downregulated proteins POLR2A, IWS1, XRCC1, TP53BP1, and KAT7 in Biological Process GO terms in combination therapy. The nodes are the GO terms

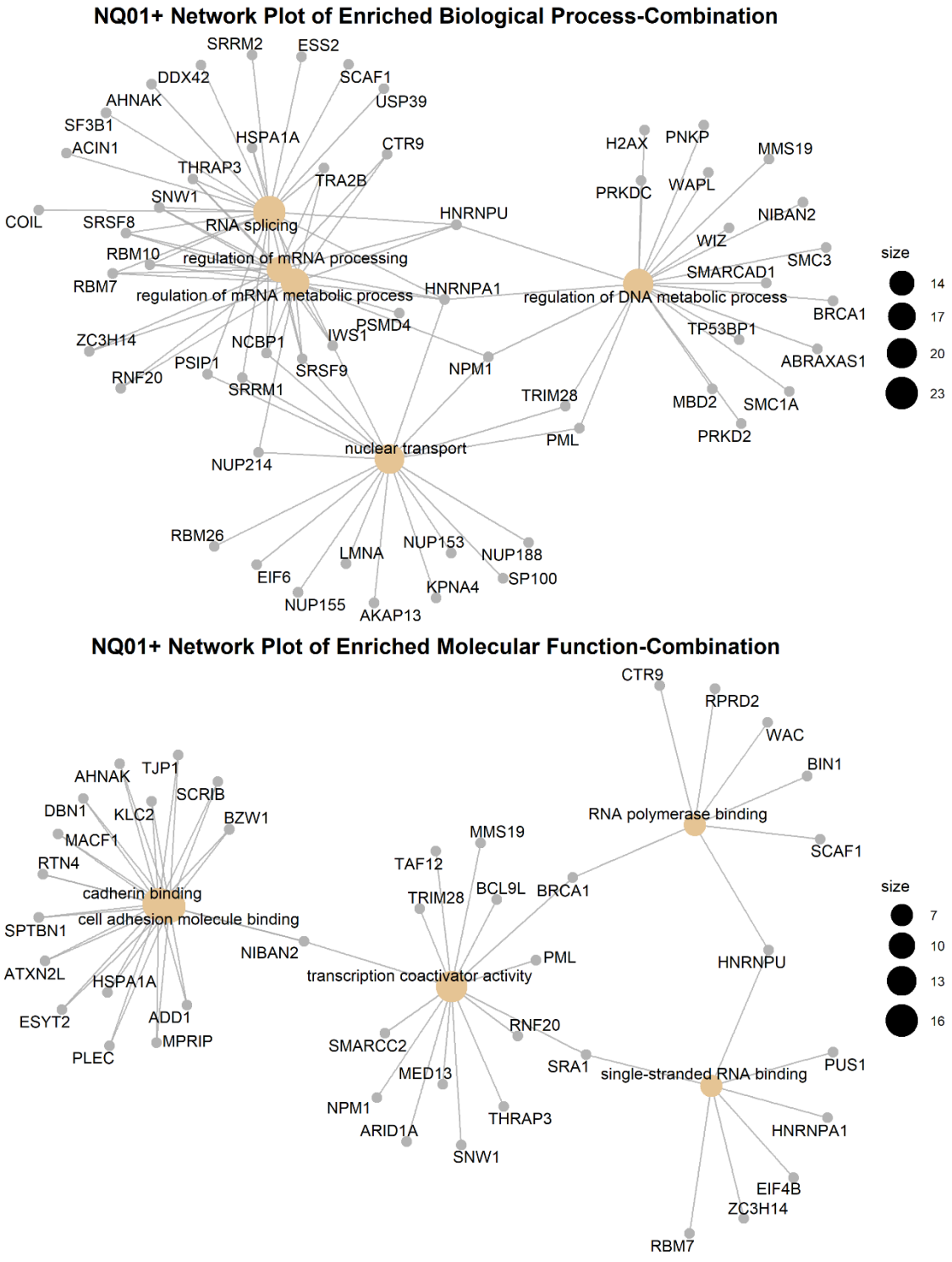


Figure 14: Gene Concept Network of upregulated proteins H2AX, HNRNPU, PSMD4, SP100 and PML in Biological Process GO terms in combination therapy. The nodes are the GO terms

Previous studies [50] have described the role of SWI-SNF complex in tumor suppression by transcriptional regulation and DNA damage response. Mutations and changes in overall expression of SWI-SNF complex subunits is associated with poor prognosis in many cancer types including breast cancers [51][52]. Upregulation of INO80B, HDAC2, PBRM1, RSF1, SMARCC2, CHD3, ARID1A, SLC9A1 in combination therapy indicates transcription suppression and DNA damage response (*Figure 19*). Mutations in PBRM1, ARID1A and SMARCC2 is associated with cancer progression and needs further investigation [50].

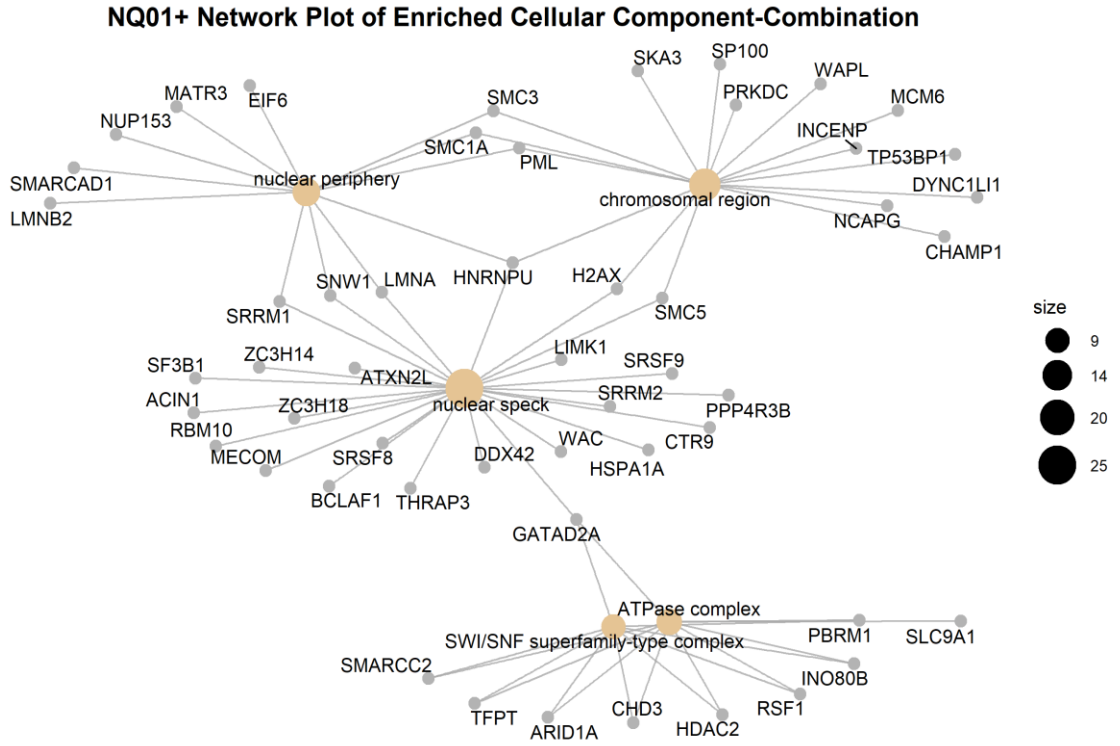


Figure 15: Gene Concept Network of upregulated proteins belonging to SWI-SNF family in combination therapy. The nodes are the GO terms

Upregulation of DNA damage sensor proteins like MMS19, FOXK1, ATM, histone deacetylation (CHD4) and transcription repression (GATAD2B) was observed in IB-DNQ

0.4 µM treated cells. MMS19 is associated with NER and HR mediated DNA repair, PNKP is associated with both non-homologous end joining (NHEJ) and base excision repair (BER) pathways, and ATM activates checkpoint signaling upon encountering DSBs and phosphorylates Ser-139 of H2AX *(Figure 16,17,18).*

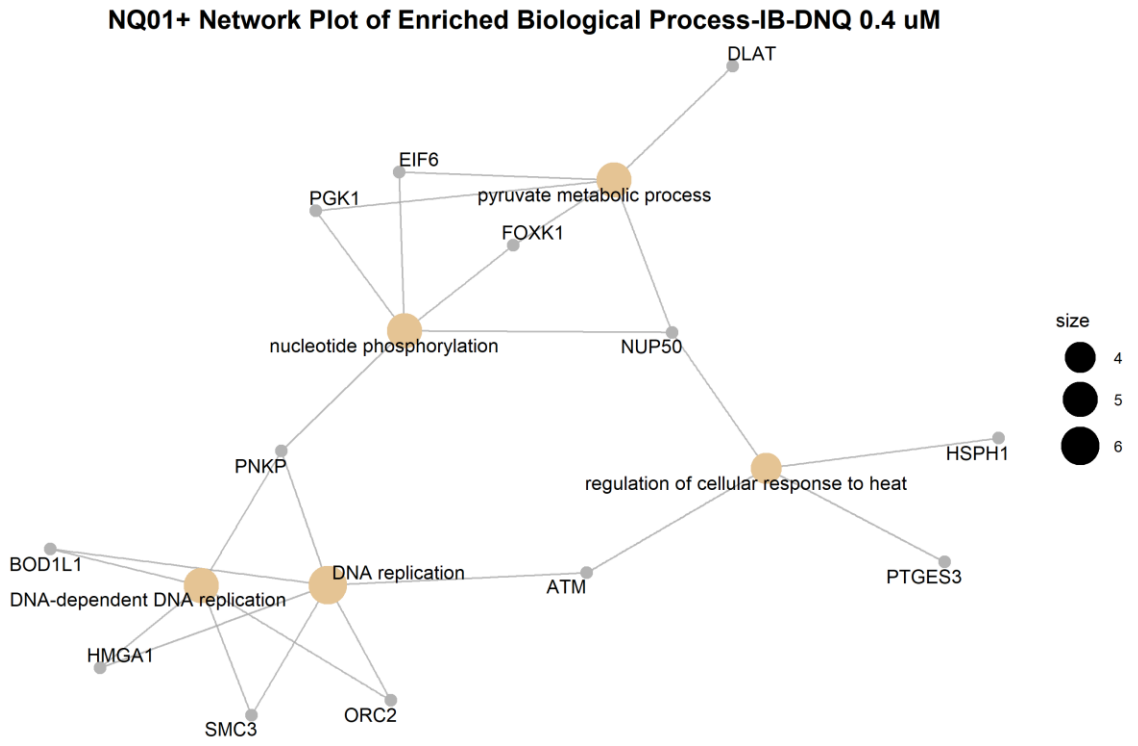


Figure 16: Gene Concept Network of upregulated proteins (PNKP, ATM, FOXK1) in Biological Pathways GO terms in IB-DNQ 0.4 µM. The nodes are the GO terms

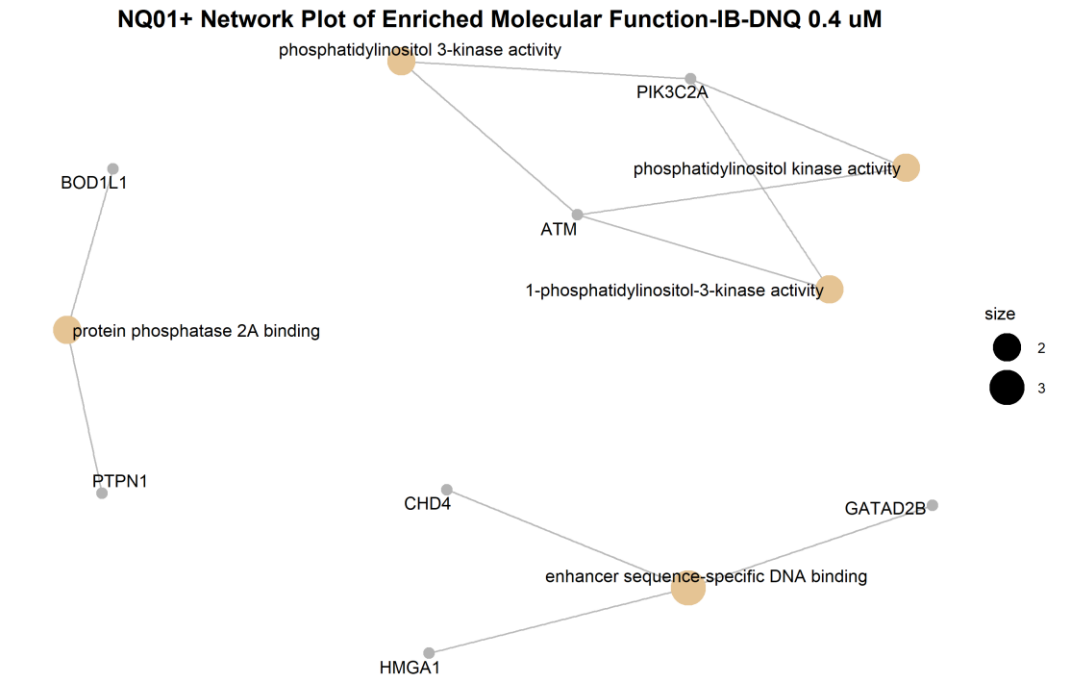


Figure 17: Gene Concept Network of upregulated proteins (CHD4, GATAD2B) in Molecular Function GO terms in IB-DNQ 0.4 µM. The nodes are the GO terms

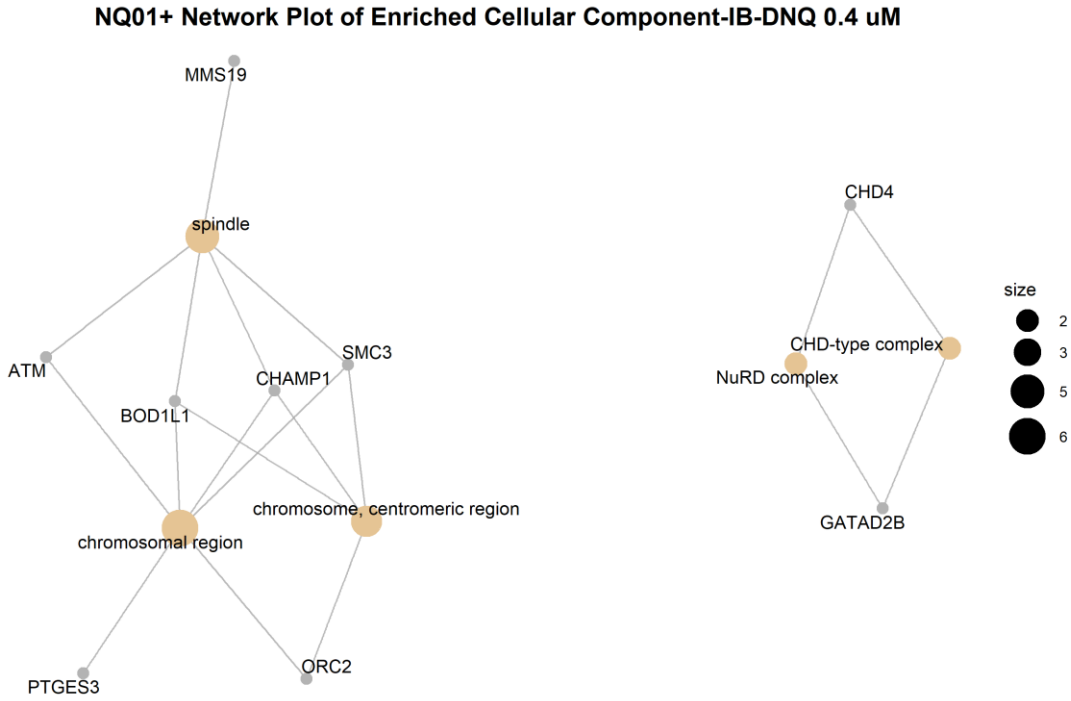


Figure 18: Gene Concept Network of upregulated proteins (MMS19, ATM, GATAD2B) in Cellular Component GO terms in IB-DNQ 0.4 µM. The nodes are the GO terms

Downregulation of most proteins related to mRNA regulation, MTORC1 pathway (TSC2, LARP), focal adhesion (CTTN) and podosome (TJP1, ASAP1, SVIL) is seen in 0.4 µM IB-DNQ. All the proteins belonging to the GTPase regulation node are associated with Ras signaling pathway such as RASAL2 inactivates Ras-cyclic AMP pathway. These proteins are also annotated in DisGeNET for their association with different forms of cancer *(Figure 19,20,21).* Podosomes are actin enriched protrusions that are responsible for the invasiveness or metastases of many cancer types including breast cancer and play a role in cell migration. They are formed in cells that need to cross tissue boundaries like monocytes, dendritic cells or macrophages. Cellular tyrosine kinases like Src and Csk play major roles in podosome regulation [53]. In cancer cells they are called invadopodia, which cause degradation of extracellular matrix by protease activity and promote crossing of tissue barrier by cancer cells. MDA-MB-231 breast carcinoma cells are well studied for podosome formation as reported by Flynn *et al.* [54].

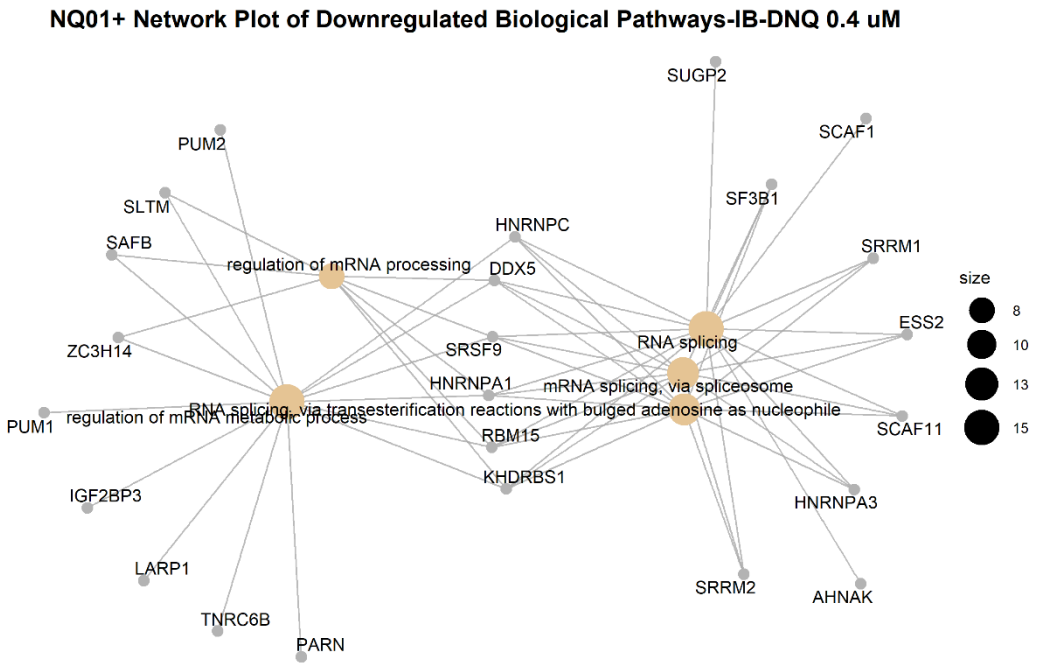


Figure 19: Gene Concept Network of downregulated proteins in Biological Pathways GO terms in IB-DNQ 0.4 µM. The nodes are the GO terms

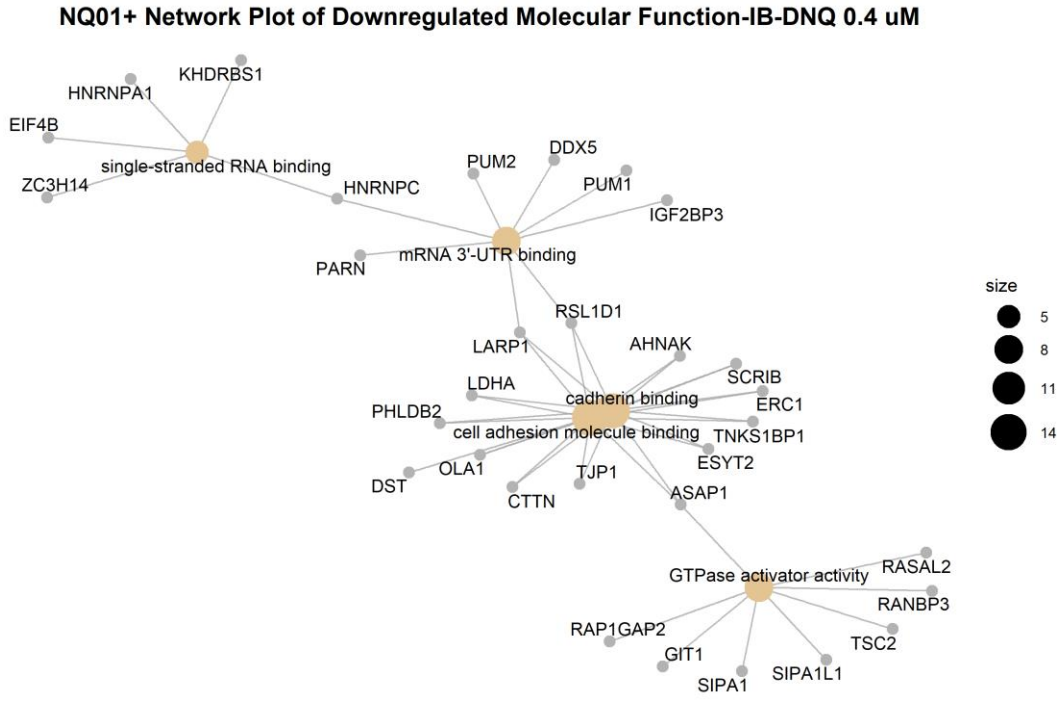


Figure 20: Gene Concept Network of downregulated proteins (TSC2, LARP1, CTTN) in Molecular Function GO terms in IB-DNQ 0.4 µM. The nodes are the GO terms

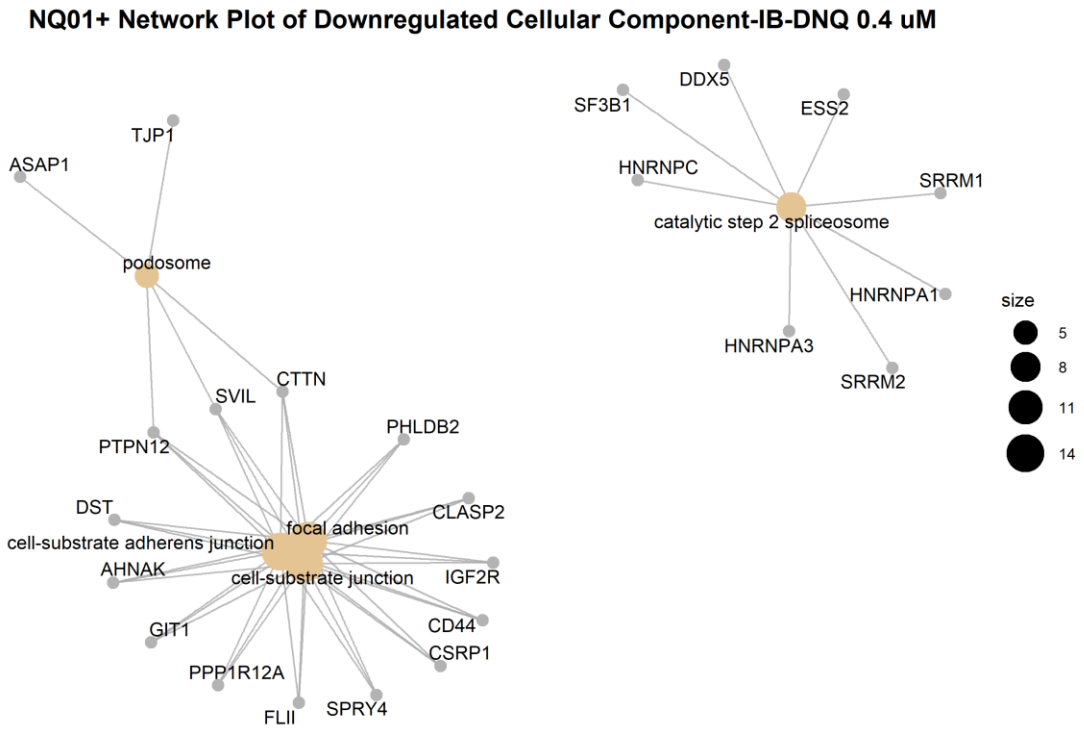


Figure 21: Gene Concept Network of downregulated proteins (TJP1, ASAP1, SVIL) in Cellular Component GO terms in IB-DNQ 0.4 µM. The nodes are the GO terms

ABL1 a tyrosine protein kinase, ATAD2 involved in proliferation and cell cycle progression in breast cancer cells, and KDM2B a histone demethylase was found to be downregulated in Rucaparib 15 µM treated cells *(Figure 22).* ABL1 is involved in DNA damage response and some of its substrates are mediators of DNA repair.

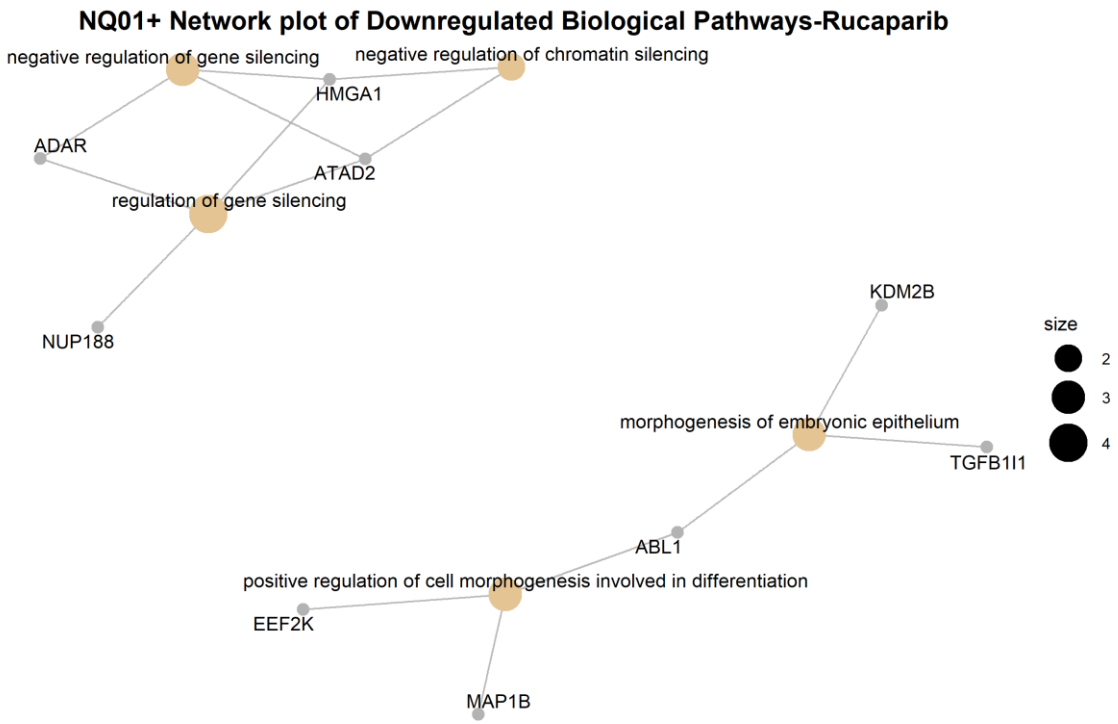


Figure 22: Gene Concept Network of downregulated proteins (ABL1, ATAD2, KDM2B) in Biological Pathways GO terms in Rucaparib 15 µM. The nodes are the GO terms

However, GO analysis have certain limitations towards extensively studied proteins and pathways. Certain areas of biology are more well studied and thoroughly annotated than others. This introduces certain bias into the statistical analysis. Also, GO is based on manual curation or computational approaches and the existing annotation databases are largely incomplete. If a protein is detected but there is no reference of it in the published literature, the algorithm will ignore that protein. The interpretation of the experiments will also change over time as more and more data become available and annotations are updated, thus affecting the reproducibility of experiments over time.

Gene Set Enrichment Analysis (GSEA)

While GO analysis uses ranked list of proteins, GSEA is threshold free and considers the whole list of proteins detected in the experiment. The hypothesis behind this method is that small changes in sets of functionally related genes or proteins may also be important in addition to genes having large expression changes. The hallmark gene set of cancer gene pathways from MSigDB is used for GSEA analysis. Upon comparing the combination therapy treated cells against all other drug treatment conditions, genes related to P53 pathway, IL2 STAT5 signaling, WNT signaling, apoptosis, G2M checkpoint, JAK STAT pathway, glycolysis, KRAS signaling and oxidative phosphorylation are found to be positively correlated in combination therapy. Whereas in the rest of the drug treated cells perturbations in MTORC1, MYC, E2F targets, DNA repair pathway genes, interferon gamma response and unfolded protein response pathways are observed (*Figure 23).* These changes were not observed in GO enrichment analysis due to the bias introduced by p value threshold with GO analysis.

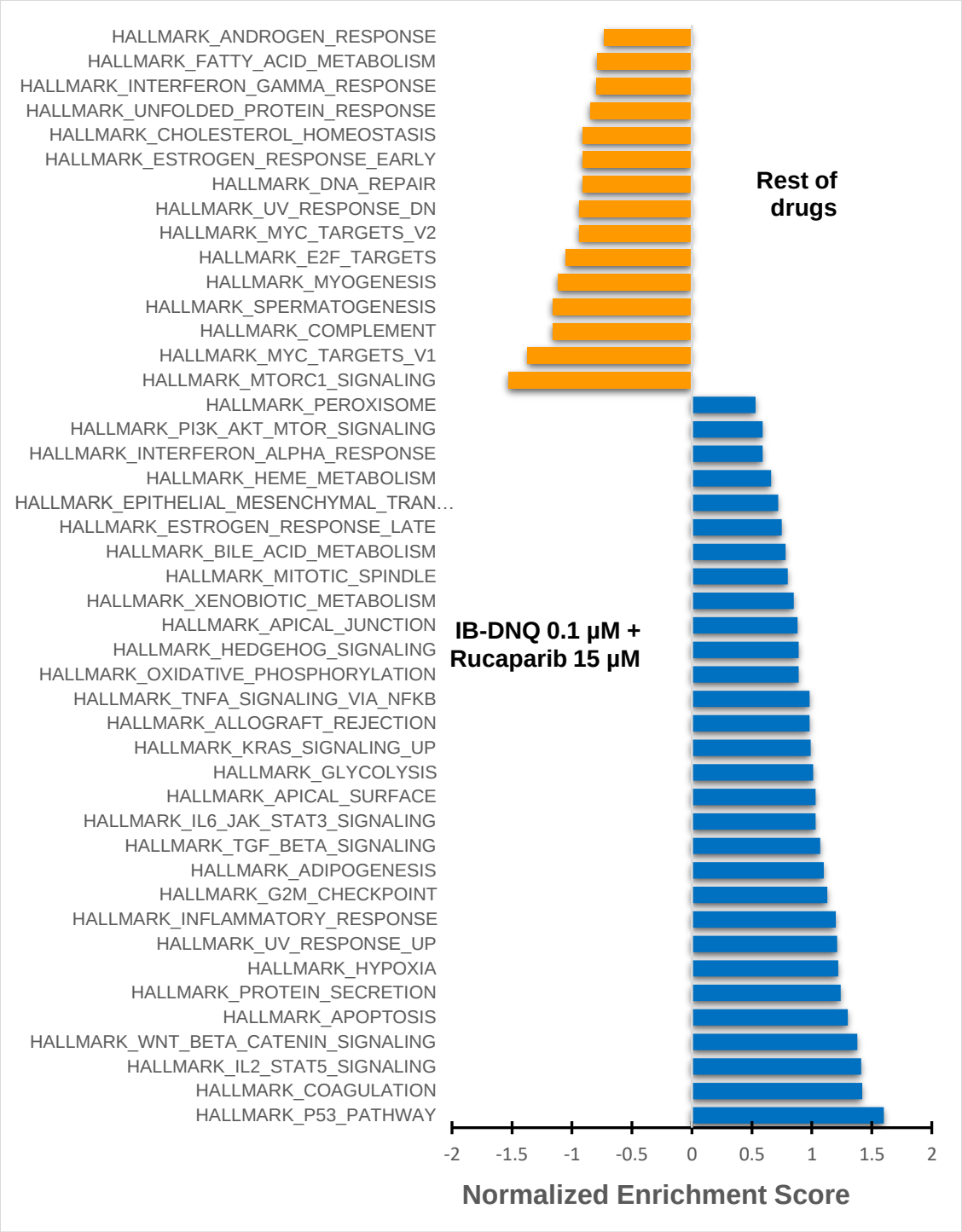


Figure 23: Barplot of GSEA analysis in combination therapy vs rest of the drugs (Rucaparib 15 µM, IB-DNQ 0.1 µM, IB-DNQ 0.4 µM)

The E2F, P53 pathway genes, G2M checkpoint genes are expressed in cells treated with combination therapy drugs, whereas genes belonging to glycolysis pathway, epithelial to mesenchymal transition (EMT), PI3K/AKT/mTOR signaling and MTORC1 signaling are expressed in single agent Rucaparib treatment (*Figure 24*). The enrichment plots of the pathways clearly show the correlation between the two conditions. As the analysis walks down the hallmark cancer gene list, it scores the genes present in the experimental gene set. The green peaks represent enrichment score, which is the maximum deviation from zero. Gene sets with a distinct peak at the beginning or end of the gene set is significant. The vertical bars in the middle represent the position of each gene towards the top or bottom of the ranked list of proteins. The bottom portion of the plot shows the value of ranking metric. A positive value indicates correlation with the first phenotype and upright peak and a negative value indicates correlation with the second phenotype and inverted peak.

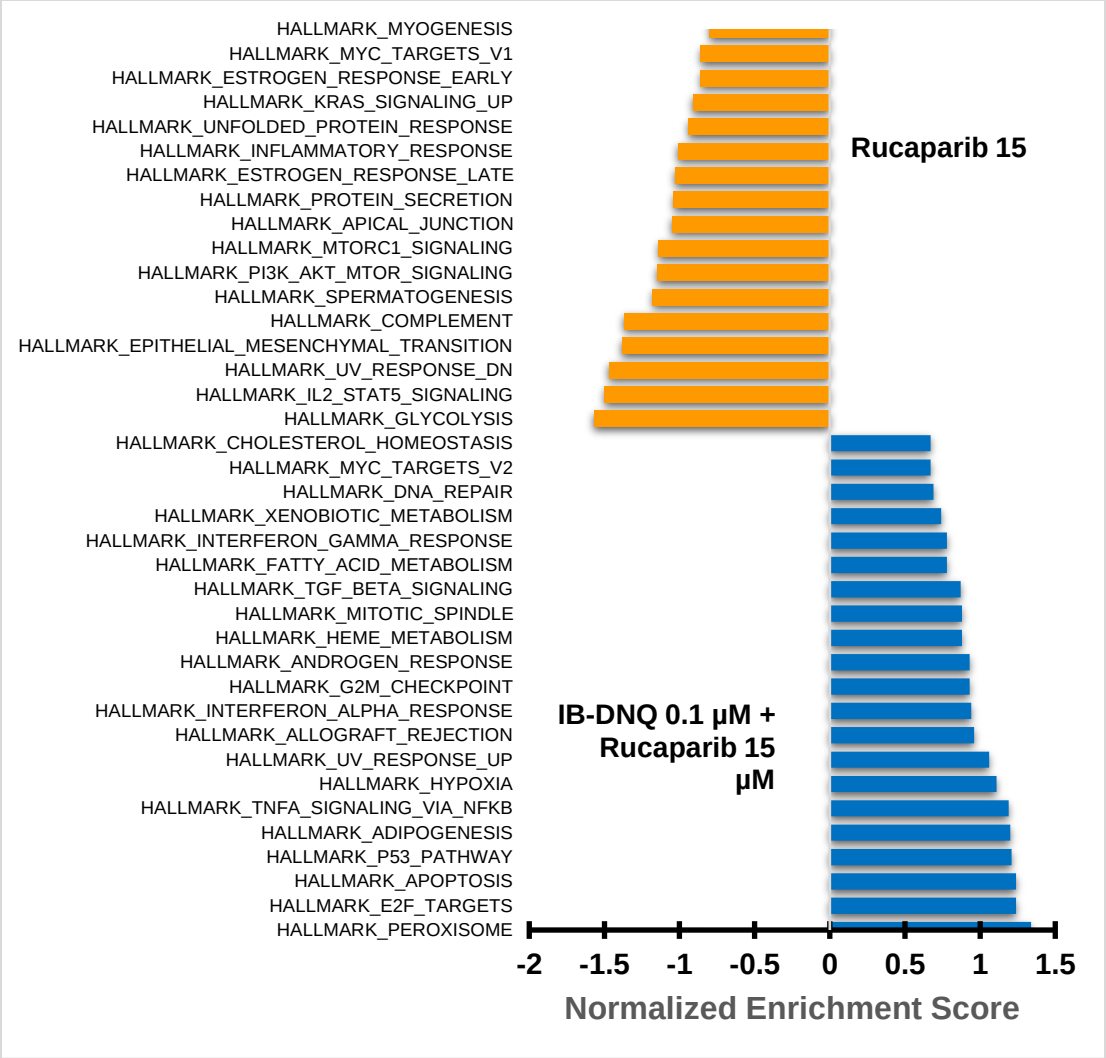


Figure 24: Barplot of GSEA analysis in combination therapy vs 15 µM Rucaparib

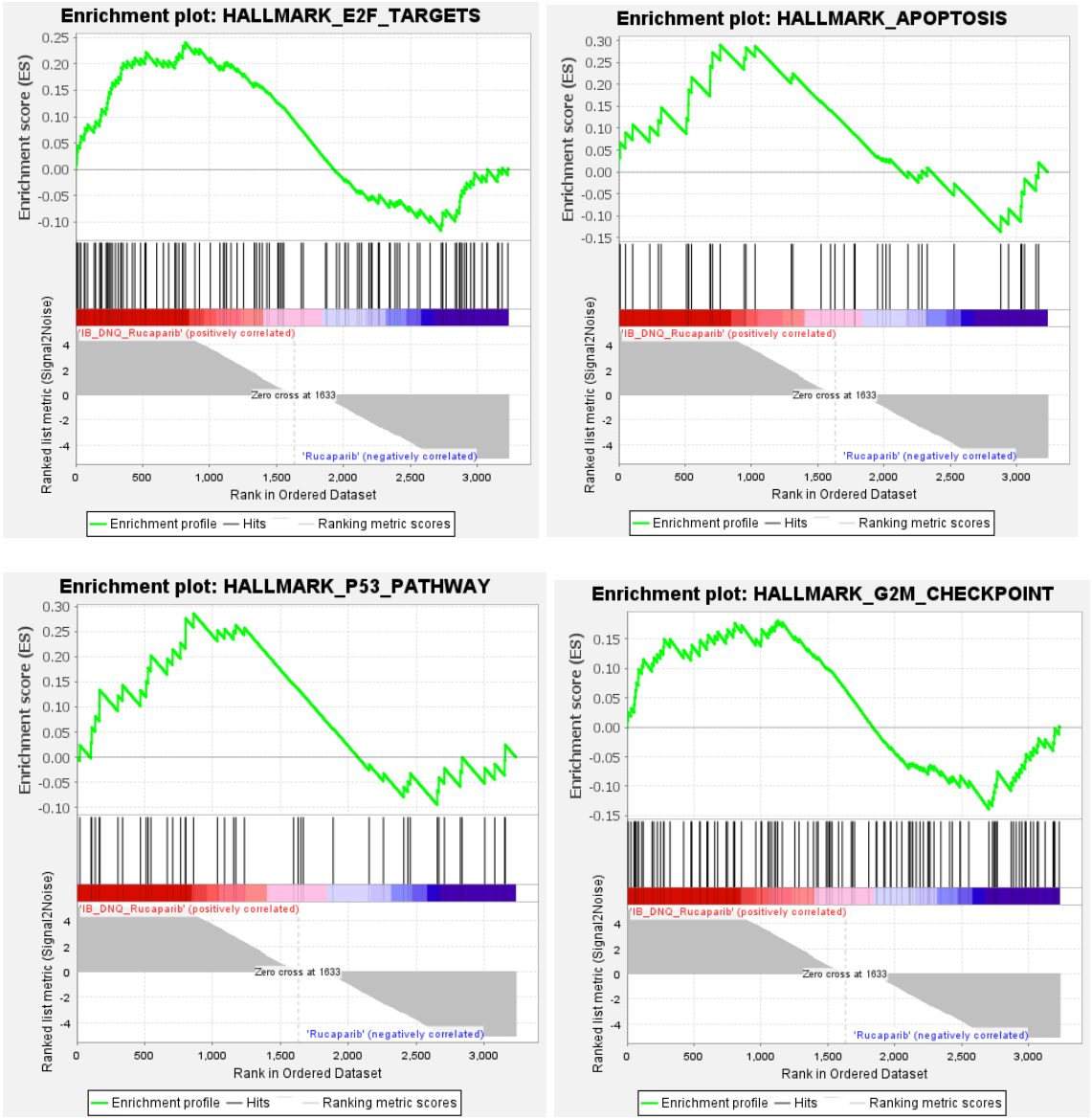


Figure 25: GSEA enrichment plots of the upregulated pathways between combination therapy and 15 µM Rucaparib. These pathways are positively correlated with combination therapy as shown by the peak at the beginning of the ranked gene list

In *Figure 25*, a comparison between drug conditions show that the genes for E2F targets, P53 pathway, G2M checkpoint and apoptosis are more enriched in the combination therapy treated cells than single agent Rucaparib as shown by positive correlation in the enrichment

plots, while genes for glycolysis, EMT, PI3K and MTORC1 signaling are more enriched in Rucaparib treatment than combination therapy (*Figure 24*).

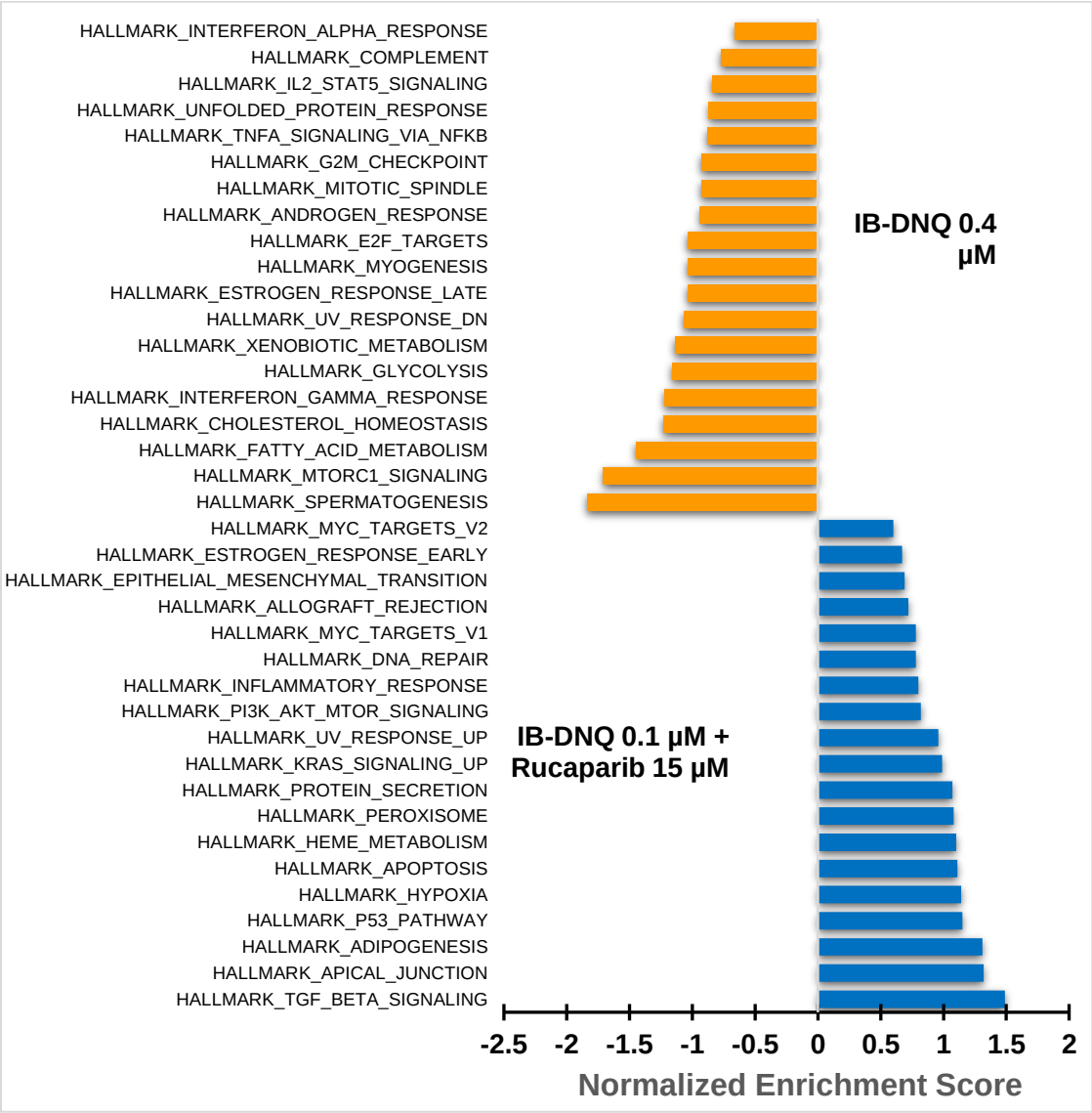


Figure 26: Barplot of GSEA analysis in combination therapy vs 0.4 µM IB-DNQ

Figures 26, 27 show the correlation of apoptosis and TGF beta signaling proteins in combination therapy compared to 0.4 µM IB-DNQ treatment, whereas proteins belonging to G2M checkpoint and E2F targets are enriched in both the treatment conditions.

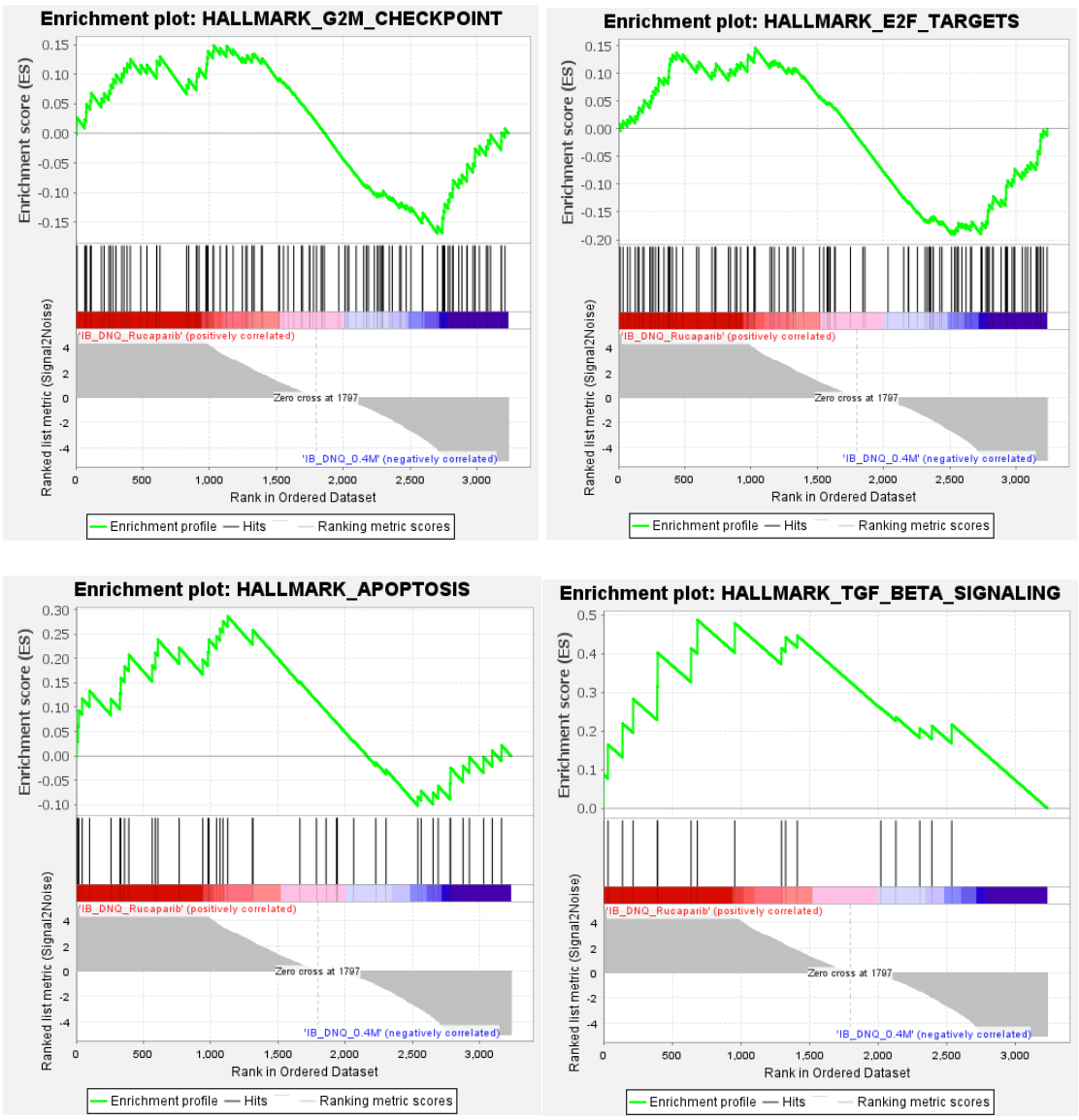


Figure 27: GSEA enrichment plots of the downregulated pathways between combination therapy and 0.4 µM IB-DNQ

Protein-Protein Interaction Network Analysis

The PPI analysis from StringDB shows 342 protein nodes and 1783 edges. To reduce the complexity of the network, the MCODE plugin (Molecular Complex Detection) in Cytoscape was used. MCODE can find clusters of highly interconnected regions in a

protein interaction network that may represent molecular complexes [55] (*Figure 28*). The network is broken down into three clusters of DNA damage related proteins, DNA repair proteins and proteins related to transcription machinery (*Figure 29*). The high interconnectivity between the proteins in a cluster represents a strong interaction amongst the proteins in biological processes. The darker color of the nodes represents higher expression based on fold change values. For example, HNRNPU blocks transcription of POLR2A transcription elongation by blocking CTD phosphorylation, H2AFX and XPC is an indication of DNA damage.

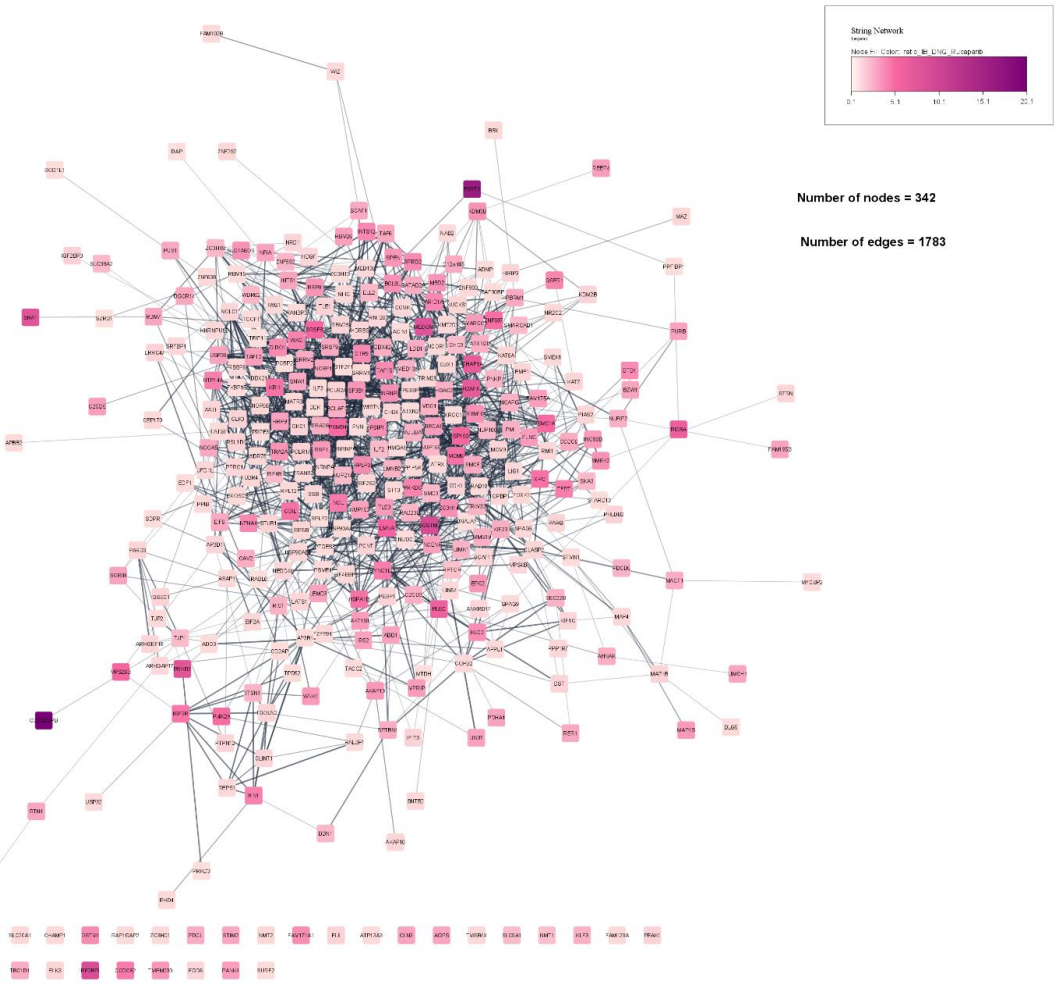


Figure 28: Protein-protein interaction network for combination therapy showing densely interconnected proteins (created using Cytoscape String app). The nodes are colored based on fold change values

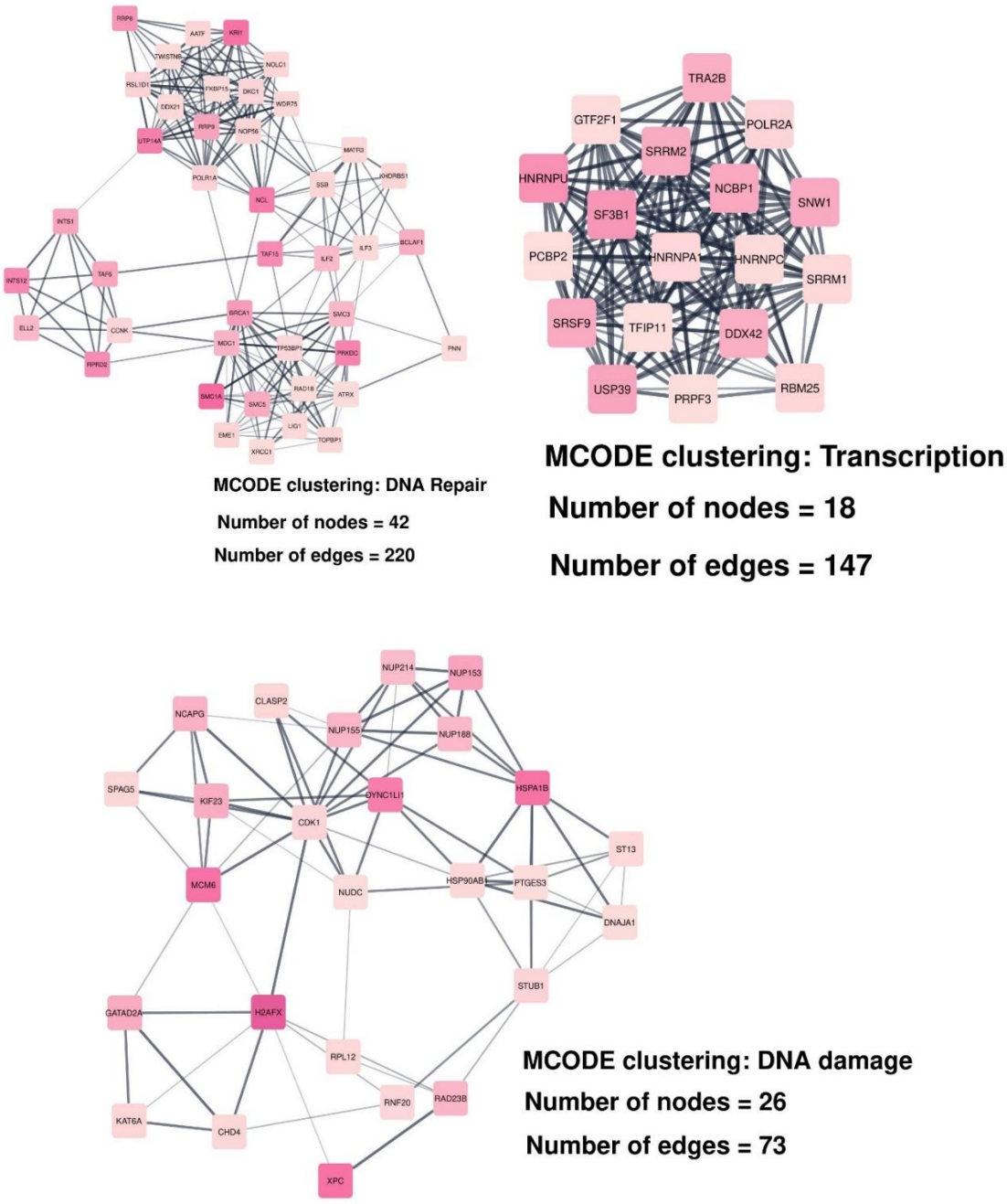


Figure 29: To reduce complexity the dense network is broken down into clusters of highly interconnected proteins from combination therapy. The clusters DNA damage, DNA repair and transcription machinery proteins may represent molecular complexes. (Created with MCODE application)

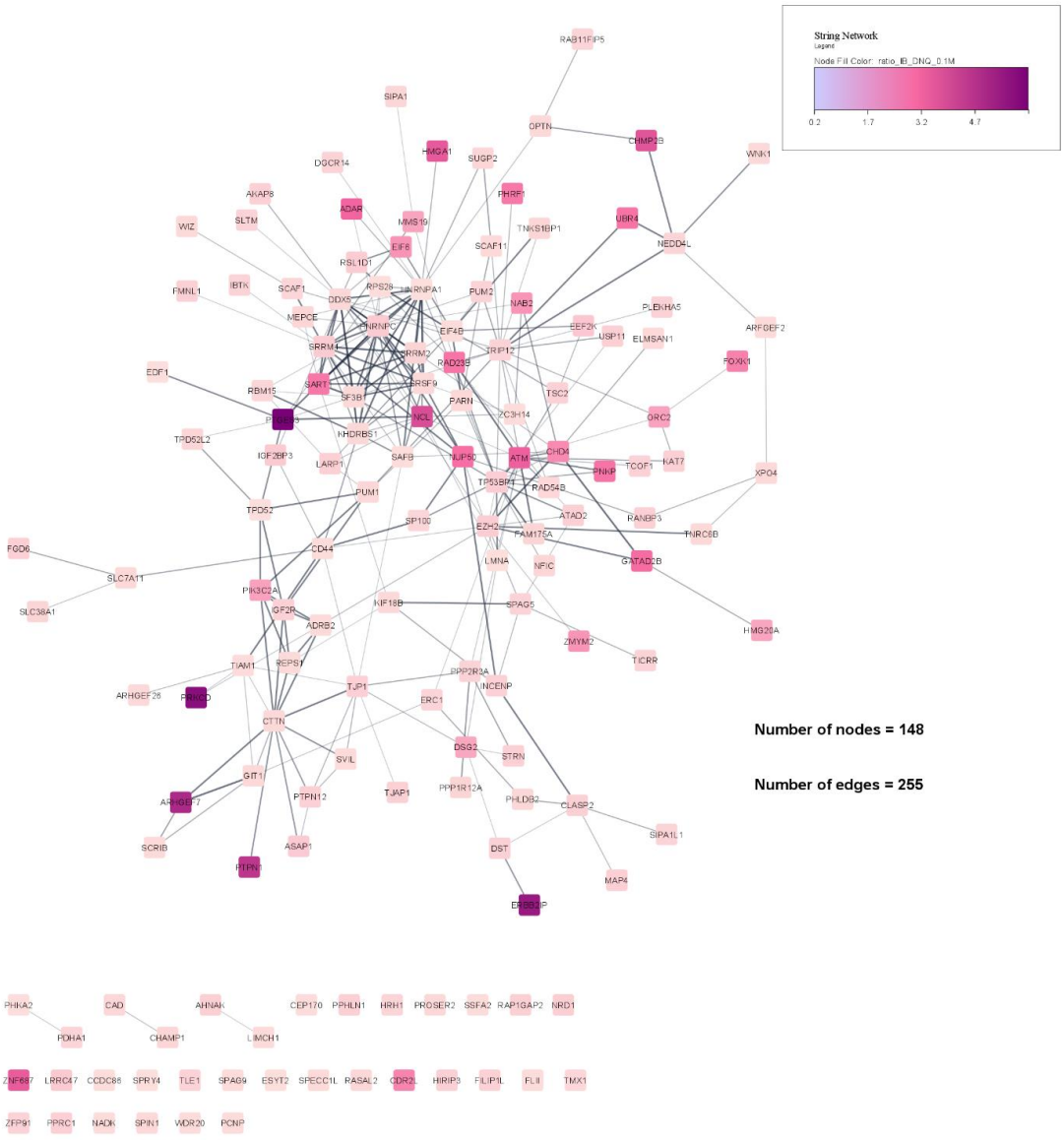


Figure 30: Protein-protein interaction network for 0.4 µM IB-DNQ (created using Cytoscape String app). The nodes are colored based on fold change values. Compared to combination therapy there are lesser interactions (edges) between proteins

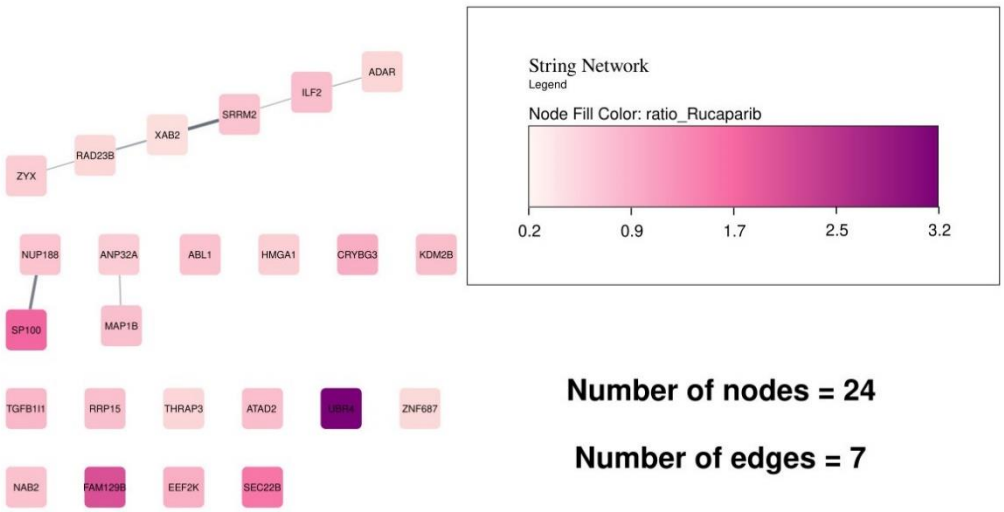


Figure 31: Protein-protein interaction network for 15 µM Rucaparib (created using Cytoscape String app). The nodes are colored based on fold change values. Very few interactions are observed between the proteins in the network

In contrast the PPI networks of 0.4 µM IB-DNQ and 15 µM Rucaparib show fewer interacting proteins with network enrichment for posttranscriptional regulation of gene expression, cell cycle process and spliceosome components (*Figure 30,31*). These findings suggest a possible structure for PPI networks with dense interconnected and self-regulated central core which has high participation in the controllability of the full network as we saw in combination therapy. The peripheral nodes are less regulated and do not play important role in the propagation of signals. However, each primary PPI database (MINT, DIP, IntAct, HPRD, BioGRID, BIND) show little overlap amongst them [56]. To increase coverage, integrating information from different primary PPI databases is required.

KSEA Results

The phosphoproteomic study presents a rich source of biological knowledge and requires novel data analysis and modeling paradigms. The next challenge is to delineate detected phosphorylation sites to their effector kinases. The KSEA provides a computational approach using literature mining to decipher the kinases responsible for the phosphorylation events. The phosphosites from the experimental dataset were used as input. After the analysis is performed, three output files are produced, KSEA kinase scores, Kinase-Substrate links and KSEA barplot. The barplot summarizes the KSEA results and only includes kinases with a substrate count cutoff of 2. The kinases that do not pass this cutoff will be excluded from the barplot and available in the KSEA kinase scores, Kinase-Substrate relationships table outputs. The p-value cutoff of 0.05 that was set before the analysis decides which kinases will be marked statistically significant. Phosphosites with decreased phosphorylation in the drug treated group will have a fold change ratio of less than 1 (where the fold change ratio was determined by *Drug/Control*), leading to $\log_2$FC value in negative and will be marked in blue. The KSEA kinase scores table provides a list of all the kinases including those that are not included in the barplot and that have at least one identified substrate in the input dataset. This may be useful for generating custom graphs for in-depth analysis. The Kinase-Substrate links table provides a list of all K-S relationships identified from the experimental phosphoproteome dataset and shows every substrate identified from the dataset that contributed to the kinase's score.

For kinases such as CDK1, CLK1 and CDK2 whose substrates that are dephosphorylated with drug, its normalized score will be negative in value (*Figure 32).* So, these kinases are deemed downregulated with combination drug treatment because its

signaling output is decreased in that group relative to control and will be marked in blue. Similarly, phosphosites with increased phosphorylation in the drug treated group will have a fold change ratio of greater than 1, leading to log_2FC value in positive and will be marked in red. Kinases such as NEK2, IKBKB, PRKCI, MAP2K5, CHUK, RPS6KA2, ATM, TGFBR2, PDK1, CHEK1, PAK2, PLK1, ROCK1, AURKC are upregulated upon combination drug treatment. Kinases with non-significant scores will be colored in black. IKBKB is a serine kinase that plays an essential role in the NF-kappa-B signaling pathway, which is activated by multiple stimuli such as DNA damage or other cellular stresses. The serine/threonine protein kinase ATM activates checkpoint signaling upon double strand breaks (DSBs) and apoptosis by acting as a DNA damage sensor and phosphorylates Ser-139 of histone variant H2AX at double strand breaks (DSBs), to regulate DNA damage response mechanism. ATM also phosphorylates BRCA1.

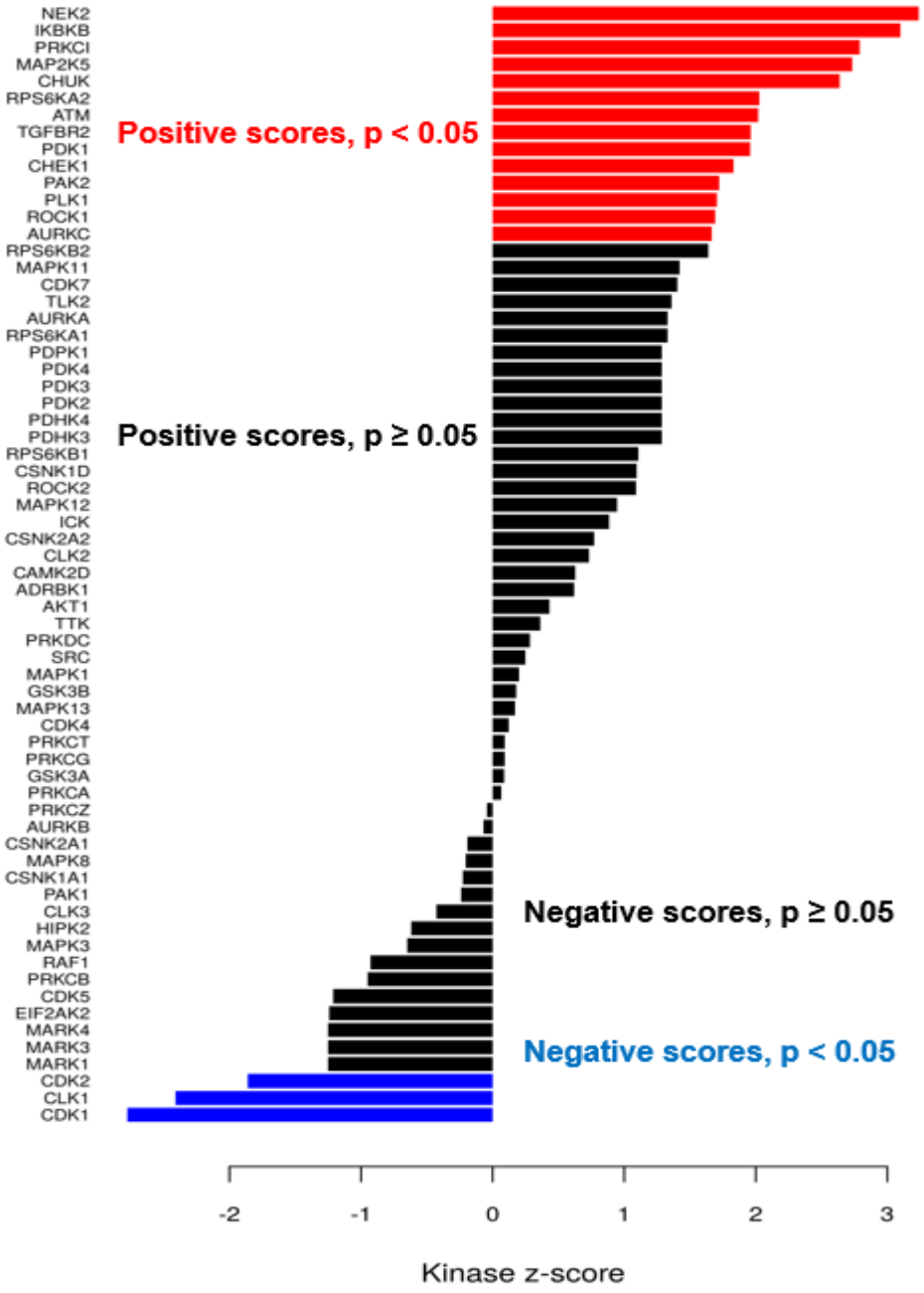


Figure 32: KSEA barplot for NQO1 positive combination treatment (15 μM Rucaparib and 0.1 μM IB-DNQ). The upregulated and downregulated kinases according to p-value are colored in red and blue respectively. CDK1, CDK2 and CLK1 are downregulated upon combination therapy

The Polo like kinase PLK1 is responsible for DNA damage response and G_2 DNA damage checkpoint recovery by inactivating the ATR/Chk1 pathway through inhibition of 53BP1 and Chk1 activator Claspin. Phosphorylation at Thr-210 is responsible for the catalytic activity of PLK1. Precise regulation of PLK1 activity is essential for cell cycle regulation but the mechanism is not fully understood. A study by Weizhe Li et. al showed that PLK1 is controlled by a balanced methylation and phosphorylation switch with the methylation occurring at Lys-209 [57]. Additional studies also suggest that phosphorylation or autophosphorylation of Ser-137 may enhance activity of PLK1 during DNA damage recovery although it is not essential for PLK1 activation [38][59]. A number of studies have revealed that *PLK1* is highly expressed in most human cancers, and its overexpression is associated with poor prognosis in cancer patients [60][61]. In breast cancer, it is associated with aggressive characteristics, such as vascular invasion, markers of proliferative activity and lack of detectable estrogen receptor [62]. There was also a close association of elevated PLK1 with triple negative tumors, considered to be poor prognosis breast cancers that generally harbor TP53 mutation. It is found that patients with *TP53* mutation and detectable PLK1 show reduced survival and are more likely to have a triple negative genotype [63]. P53-null cells are unable to down-regulate PLK1 levels in response to clinically-relevant genotoxic drugs. Cancer cells sometimes show increased dependence on normal cellular functions of certain genes which are not classical oncogenes, such as PLK1. This phenomenon is called non-oncogene addiction. In a recent genome-scale shRNA (short hairpin RNA) screen of the human breast cancer, PLK1 was a hit in several TNBC cell lines, indicating its importance for growth and survival of these breast cancer cells [64]. PLK1 has drawn a lot of attention because its overexpression is tightly associated

with neoplastic transformation of human cells. PLK1 overexpression is thought to promote tumorigenesis by overriding cellular checkpoints and inducing genetic instability. Reversing oncogene addictions in cancer cells has been shown to induce apoptotic cell death. Therefore, increased sensitivity of cancer cells to PLK1 interrogation may likely stem from their altered signaling pathways and biochemical steps in PLK1-addicted cancer cells. Some contrasting studies show a tumor-suppressive role for Plk1 in certain tumors [65].

PLK1 is a CDK1 kinase but in the KSEA analysis for combination therapy drugs it is in the upregulated area whereas CDK1 is downregulated upon drug treatment. This conflict is explained by the phosphorylation of CDK1 on Tyr-15 by WEE1. PLK1 plays a central role in the G_2/M transition of the mitotic cell cycle by phosphorylating CCNB1, CDC25C, FOXM1, CENPU, PKMYT1/MYT1, PPP1R12A/MYPT1 and WEE1. It is part of a regulatory circuit that promotes the activation of CDK1 by phosphorylating the positive regulator CDC25C and inhibiting the negative regulators WEE1 and PKMYT1/MYT1 [66]. WEE1 acts as a negative regulator of entry into mitosis (G_2 to M transition) by protecting the nucleus from cytoplasmically activated cyclin B1-complexed CDK1 before the onset of mitosis by mediating phosphorylation of CDK1 on Tyr-15. Phosphorylation of CDK1 at Tyr-15 by WEE1 and PKMYT1 reduces kinase activity and is found in our combination treatment data (*Table 4*). WEE1 specifically phosphorylates and inactivates cyclin B1-complexed CDK1, reaching a maximum during S and G_2 phase and decreases at M phase when it is hyperphosphorylated. The inability of PLK1 to inhibit WEE1 can be one reason for the downregulation of CDK1 [67].

Kinase.Gene	Substrate.Gene	Substrate.Mod	Source	log2FC
WEE1	CDK1	Y15	PhosphoSitePlus	-0.89164
PKMYT1	CDK1	Y15	PhosphoSitePlus	-0.89164

Table 4: CDK1 is phosphorylated on Tyr-15 by WEE1 and PKMYT1

This also explains the CDK1 and CDK2 alteration following IB-DNQ treatment. IB-DNQ undergoes a futile redox cycling in NQO1 positive cells through a twostep process generating two superoxide moieties, which generate massive amounts of H_2O_2 that carries out DNA damage. The damaged DNA causes cell cycle arrest by recruiting ATM and ATR. The ATR is specifically recruited to single-stranded DNA breaks (SSBs) and coats the single-stranded DNA with replication protein A (RPA) and phosphorylates Chk1. ATM is recruited to double-stranded DNA breaks (DSBs) by MRN (Mre11-Rad50-Nbs1) complex and phosphorylates Chk2 [68]. Upon activation, Chk1 and Chk2 in turn phosphorylate WEE1 and antagonize the function of Cdc25 phosphatases, which leads to accumulation of inhibitory phosphates on Tyr-15 and Thr-14. WEE1 phosphorylates CDK1 on Tyr15, resulting in CDK1 inhibition and a stalled cell cycle [69]. Since CDK1 plays a regulatory role in homologous recombination mediated DNA repair pathway by phosphorylating BRCA1, its inhibition stops DSB repair in BRCA1 positive as well as BRCA mutated cells [70]. This causes increased hypersensitivity of cancer cells to PARP1 inhibitors like Rucaparib. Also, since CDK1/2 phosphorylates MRN complex, CtIP, Ku, PIN1, RPA, Exo1, and DNA2, its inhibition compromises single-stand DNA end resection in HR repair pathway [71]. The combined action of CDK1/2 inhibition by reactive oxygen species

(ROS) generation by IB-DNQ and PARP inhibitor Rucaparib prevented repair of DNA break and ultimately death of cancer cells [72] (*Figure 33*).

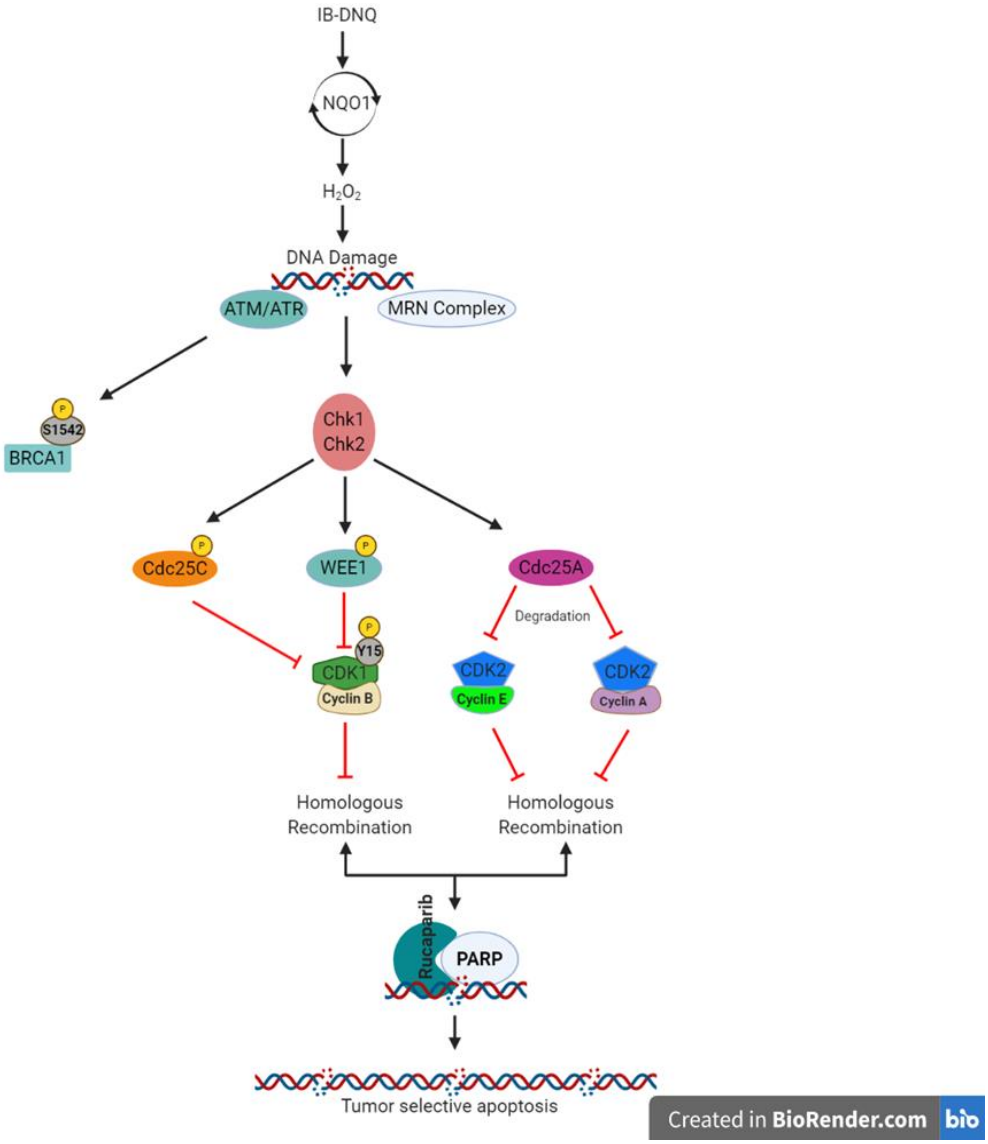


Figure 33: Schematic of CDK downregulation upon IB-DNQ treatment

Overrepresented sequence patterns among detected phosphorylation sites may correspond to phosphorylation motifs of kinases. This strategy is useful for detecting novel phosphorylation motifs of uncharacterized kinases. PLK1 (Polo-like kinase) contains PBD

(Polo Box Domain) that binds to substrates containing the consensus sequence S-pS/pT-P/X (*Figure 34*), where pS and pT stand for phosphorylated serine and threonine respectively. This is similar to the consensus sequence of CDK1 phosphorylation (*Figure 35*).

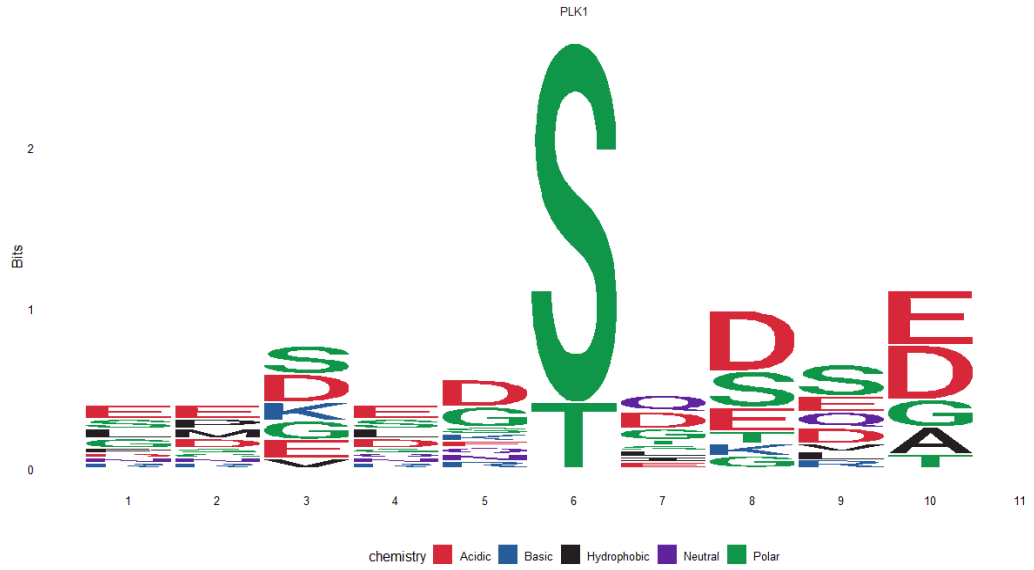


Figure 34: PLK1 consensus motif (created with ggseqlogo R package)

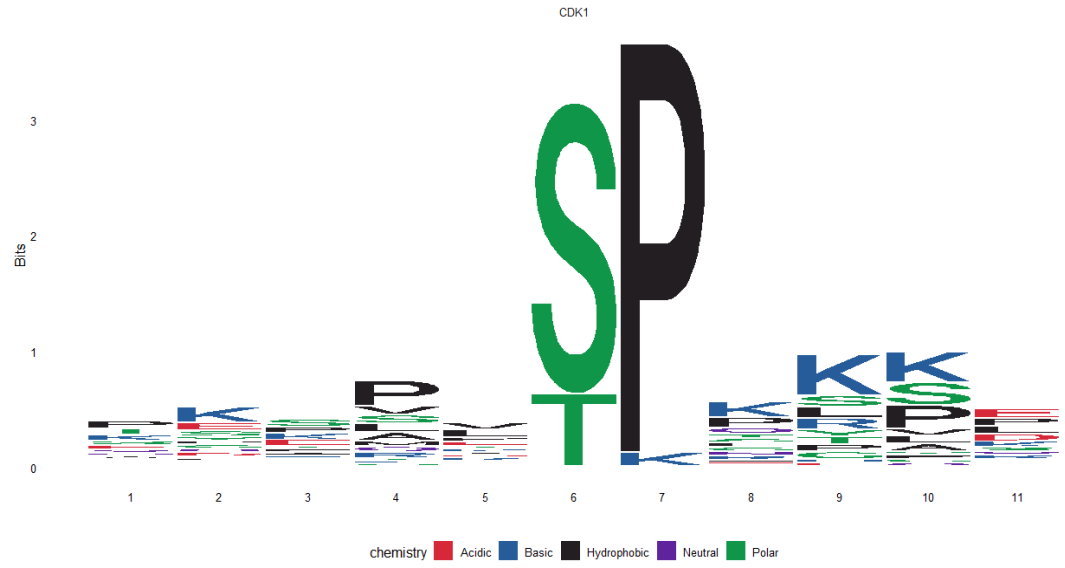


Figure 35: CDK1 consensus motif (created with ggseqlogo R package)

It is hypothesized that CDK1 primes a substrate for further phosphorylation and regulation by PLK1. In the absence of substrate, PBD binds to the kinase domain of PLK1 and inhibits its activity. After PBD binds to a phosphorylated substrate, it will move out of the active site of PLK1, which allows PLK1 to phosphorylate the substrate at a second site. This ensures that free kinase is held in an inactive state until it binds the correct substrate [11]. So, in the absence of CDK1 or downregulation of CDK1 in combination treatment, PLK1 is held in an inactive state by the PBD which explains its presence in the unregulated area.

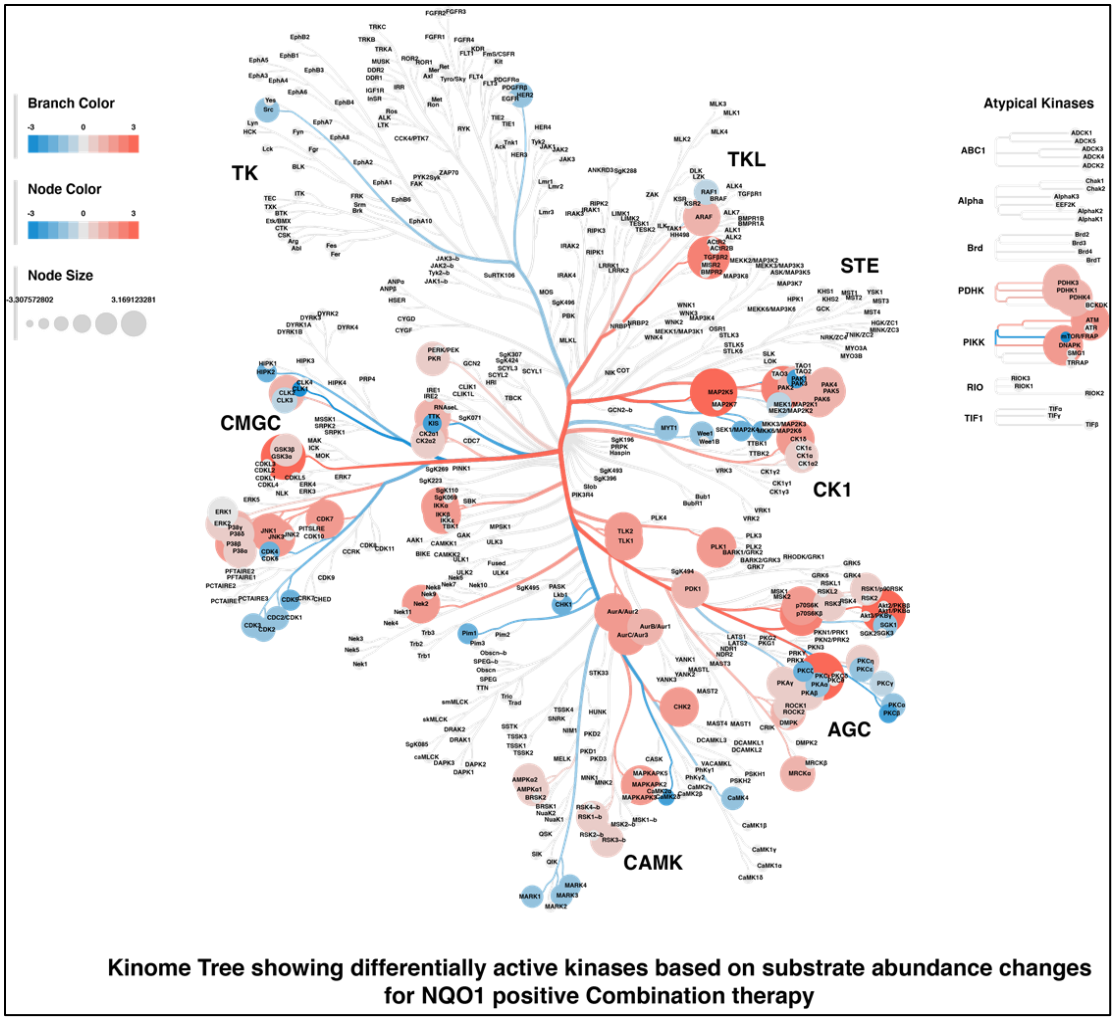


Figure 36: Kinome tree showing differentially active kinases in 0.1µM IB-DNQ + 15µM Rucaparib combination treatment (as determined by substrate abundance)

Visualization of the quantitative data from the KSEA substrate analysis on a kinome tree suggests more changes in overall kinome activity with most of the upregulated kinases belonging to the AGC, CMGC and CK1 family but none in the TK family (*Figure 36*).

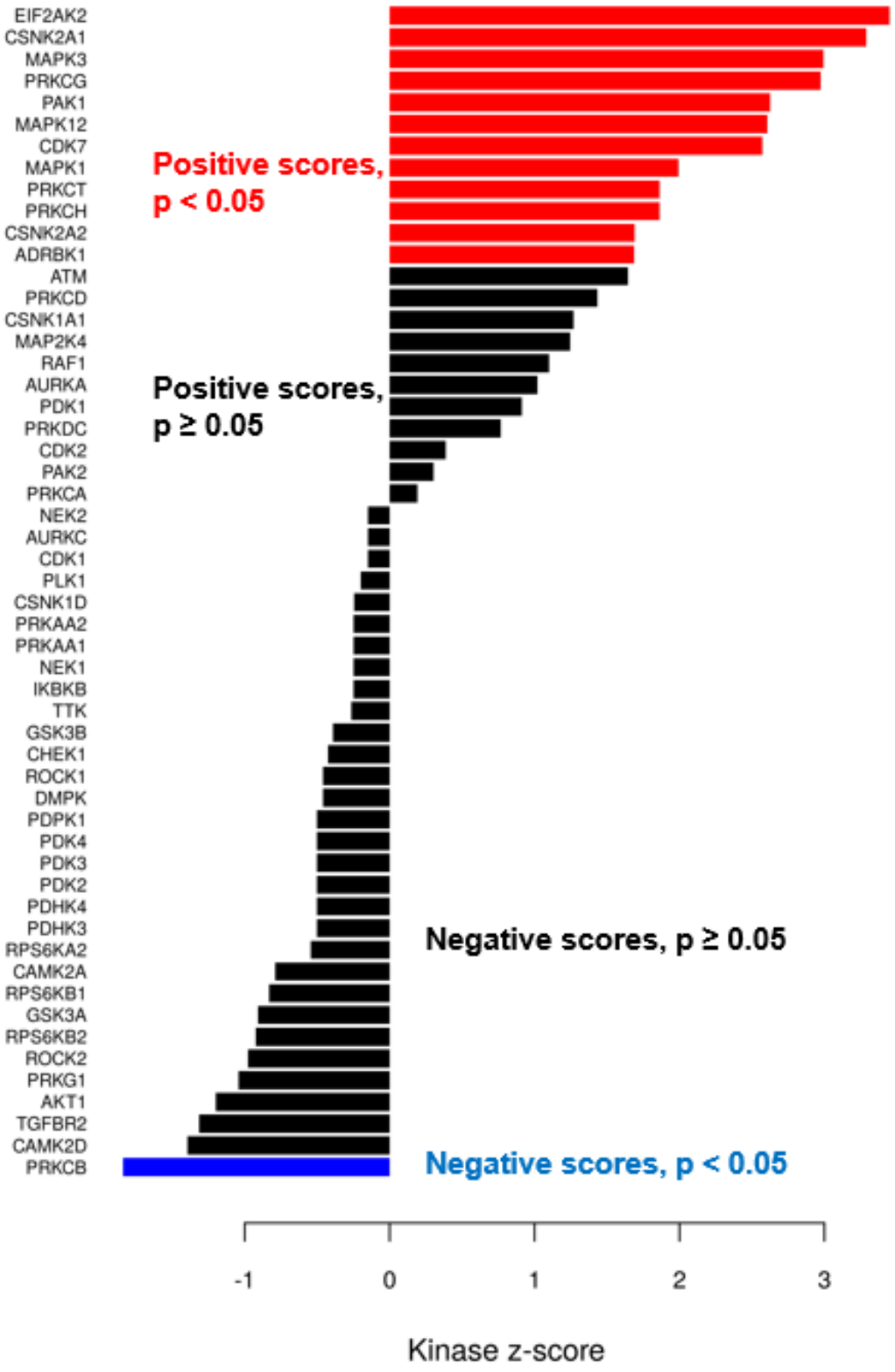


Figure 37: KSEA barplot for NQO1 positive 0.4 μM IB-DNQ treatment

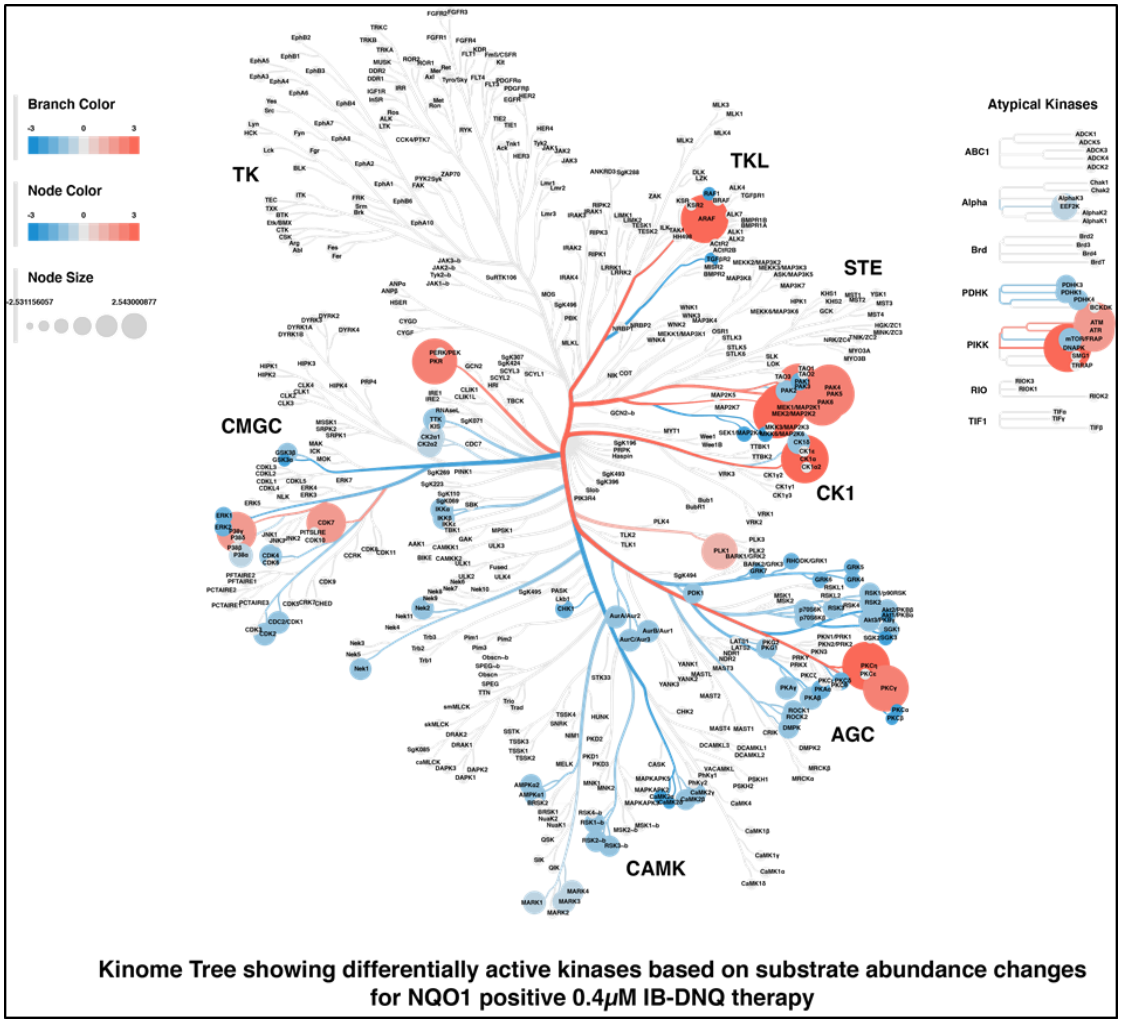


Figure 38: Kinome tree showing differentially active kinases in 0.4μM IB-DNQ (as determined by substrate abundance)

More changes in kinase activity were observed in the combination therapy and 0.4 μM IB-DNQ as compared to the low dose 0.1 μM IB-DNQ and single agent Rucaparib 15 μM treatment (*Figure 37,39,41*). Visualization of the quantitative data on the kinome tree shows decreases in kinase activity in all kinase families in Rucaparib 15 μM treatment while in IB-DNQ treatments both increases and decreases are observed in AGC and CK1

family kinases (*Figure 38,40,42*). This information can be useful to understand the interrelationships between kinases and develop therapeutic strategy targeting kinases.

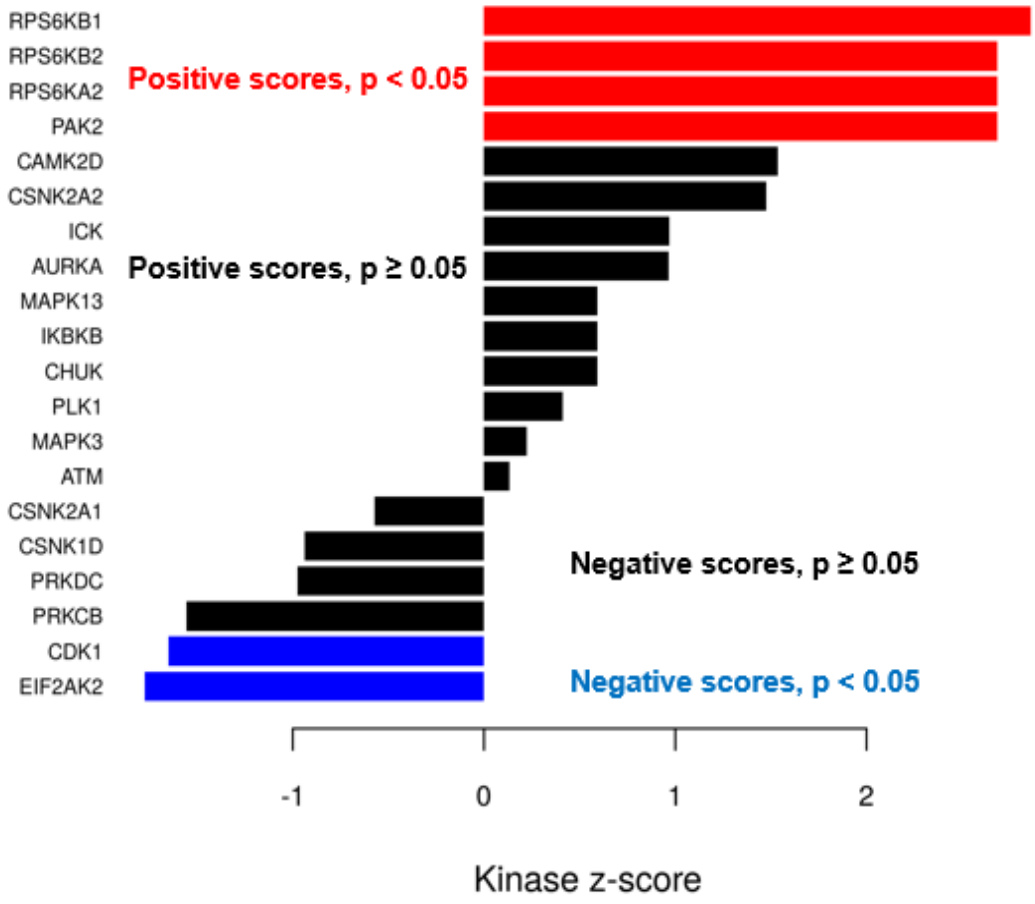


Figure 39: KSEA barplot for NQO1 positive 0.1 μM IB-DNQ treatment

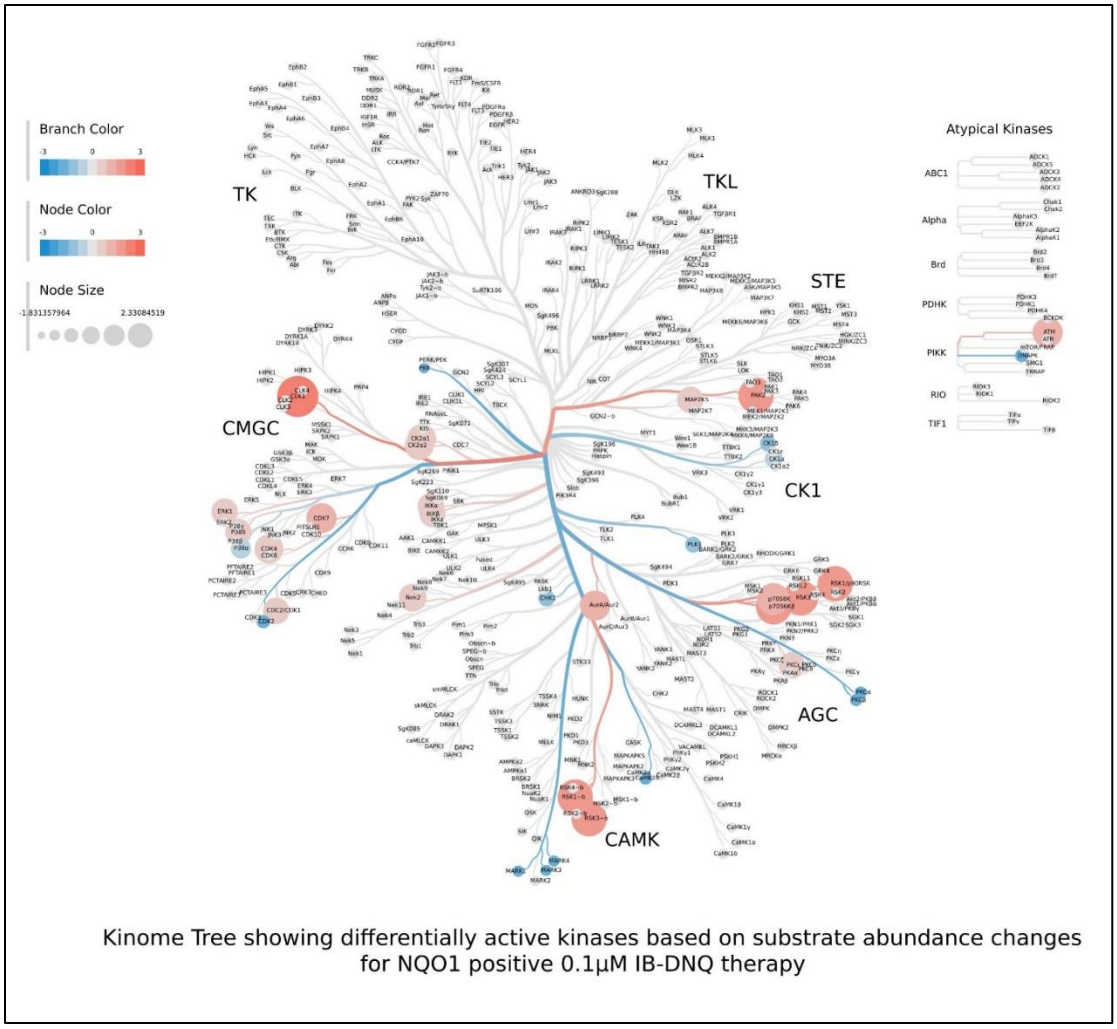


Figure 40: Kinome tree showing differentially active kinases in 0.1µM IB-DNQ (as determined by substrate abundance)

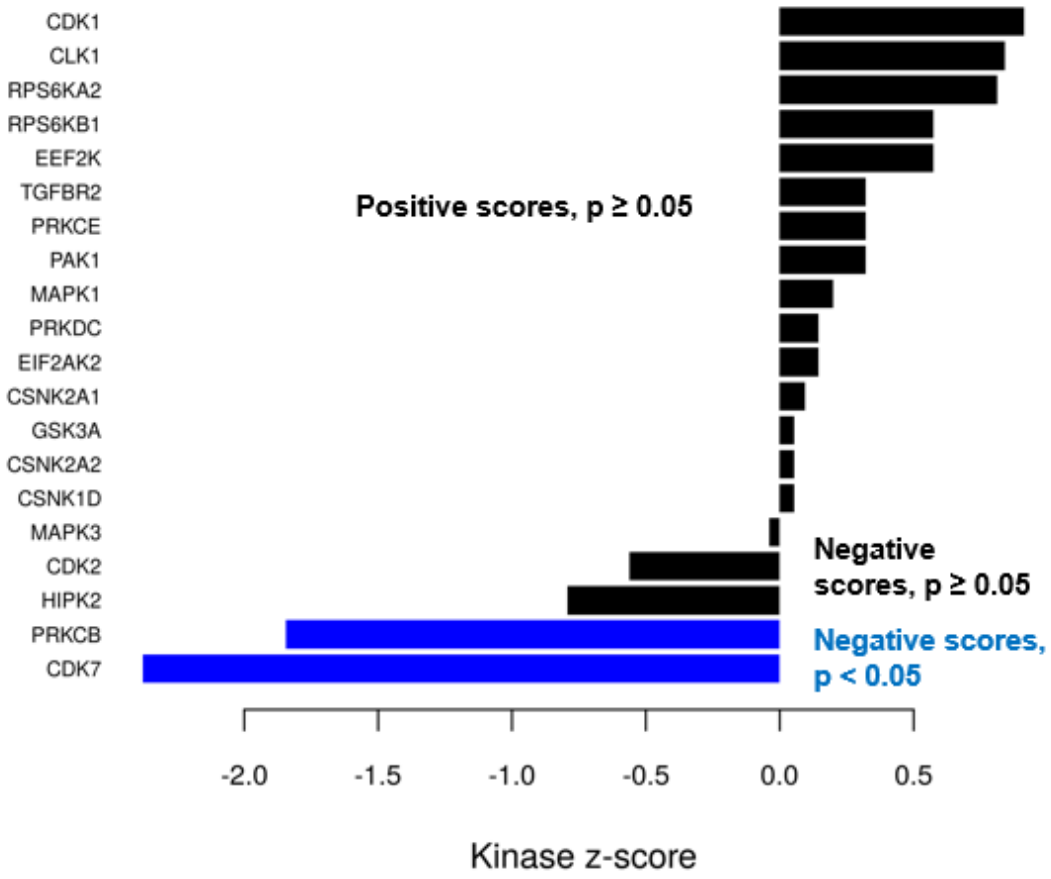


Figure 41: KSEA barplot for NQO1 positive 15 μM Rucaparib treatment

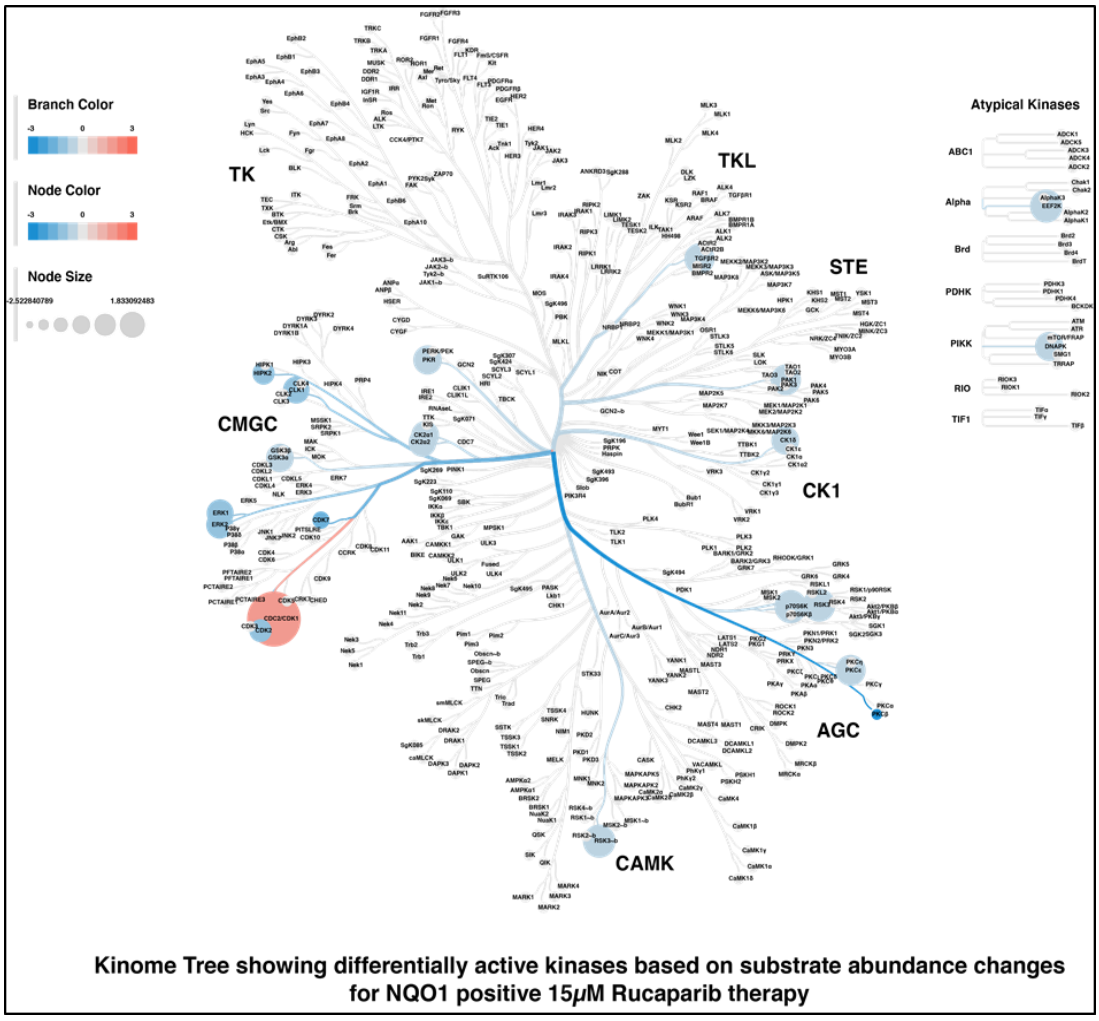


Figure 42: Kinome tree showing differentially active kinases in 15μM Rucaparib single agent treatment (as determined by substrate abundance)

Kinase Activation and Substrate Phosphorylation

The detection of the phosphorylation sites and their dynamics under different cellular conditions has necessitated the reconstruction of transient kinase-substrate interaction networks, that are essential for mechanistic understanding of cellular behavior and therapeutic intervention by computational analysis of underlying data. The data from KSEA analysis was used for network analysis of kinase-substrate interactions. The red

triangles represent kinases, the green diamonds represent transcription factors, and the blue circles represent all other substrates. The hub proteins or proteins with high interconnectivity with other proteins can have large downstream signaling effects upon perturbation. CDK1, CDK2, MAPK3 and PRKCB are highly interconnected kinase in kinase interaction data from combination therapy. For example, CDK1 can trigger hundreds of time resolved downstream signaling events through independent phosphorylations (*Figure 43*).

Dense interconnectivity between the kinases is observed in the kinase-substrate interaction network of 0.4 µM IB-DNQ (*Figure 44*). The substrates ate largely located towards the periphery of the network suggesting the kinases are interacting amongst themselves. In single agent Rucaparib, very few kinase-substrate interactions are visible (*Figure 45)*. Based on these interaction networks, it is possible to predict the clinical outcomes of the drugs used in triple negative breast cancer by examining changes in connectivity of the hub proteins in tumor cells.

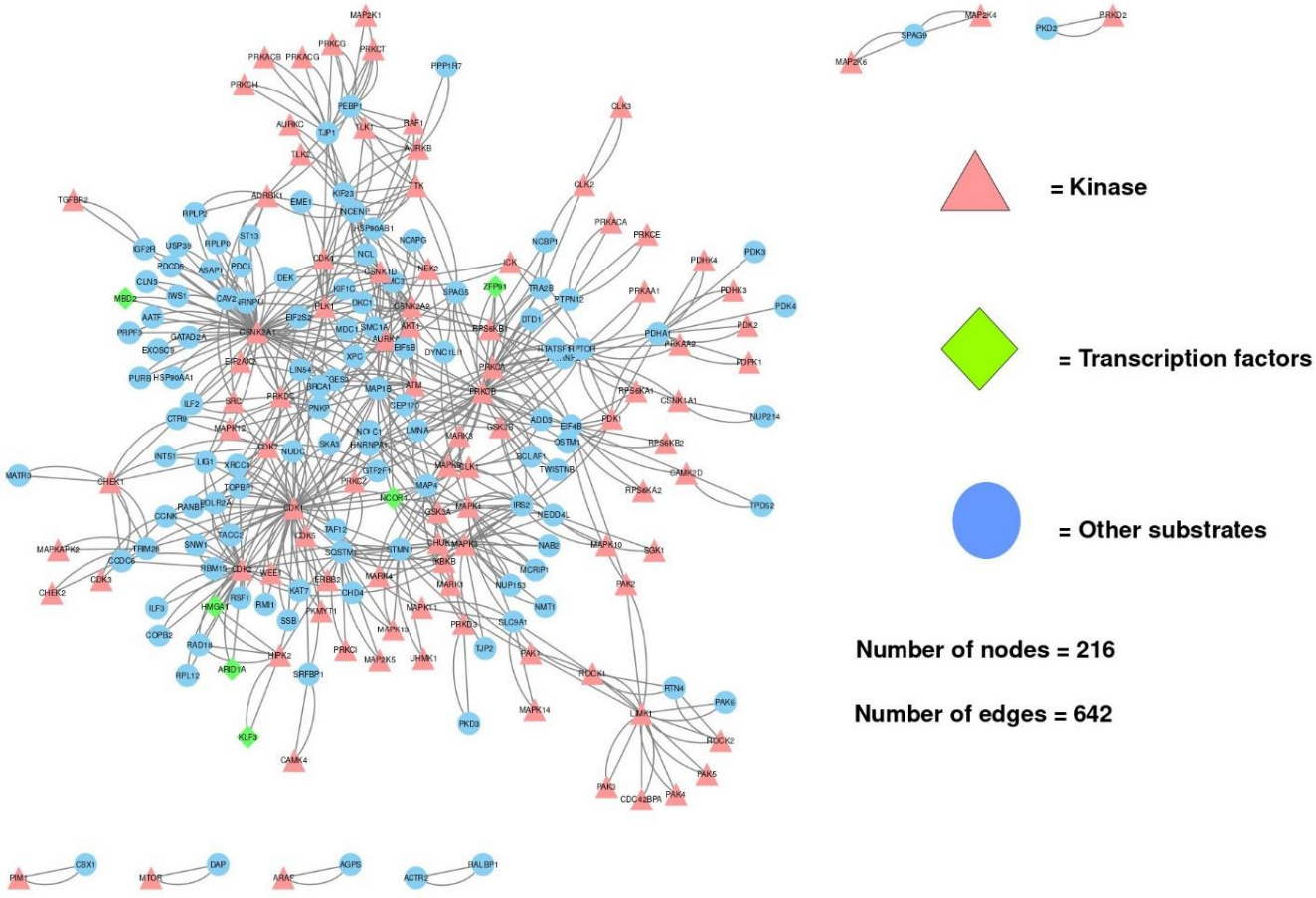


Figure 43: Kinase-Substrate Interaction Network in Combination therapy (created with Cytoscape). Certain kinases like CDK1 are densely connected to other substrates

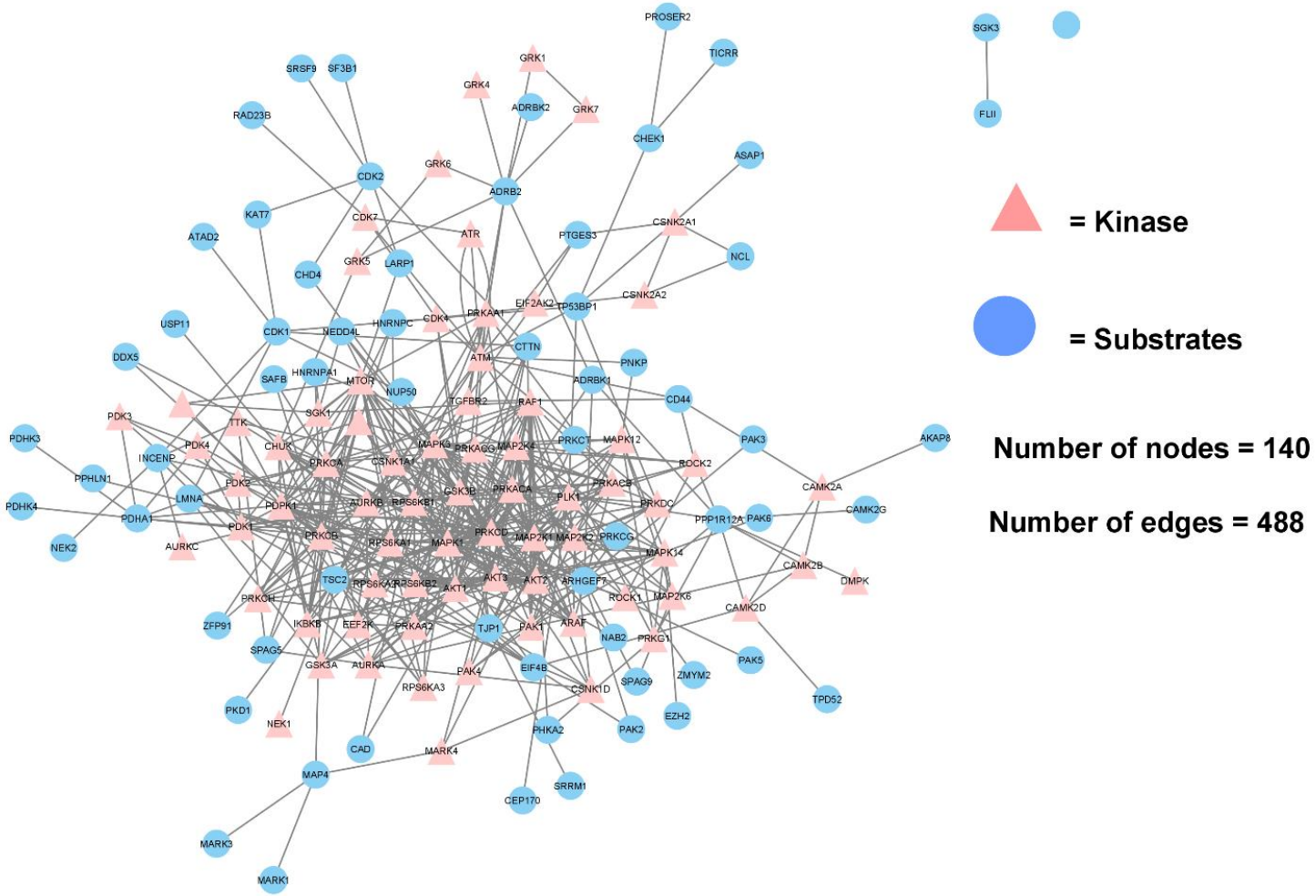


Figure 44: Kinase-Substrate Interaction Network – 0.4 µM IB-DNQ (created with Cytoscape)

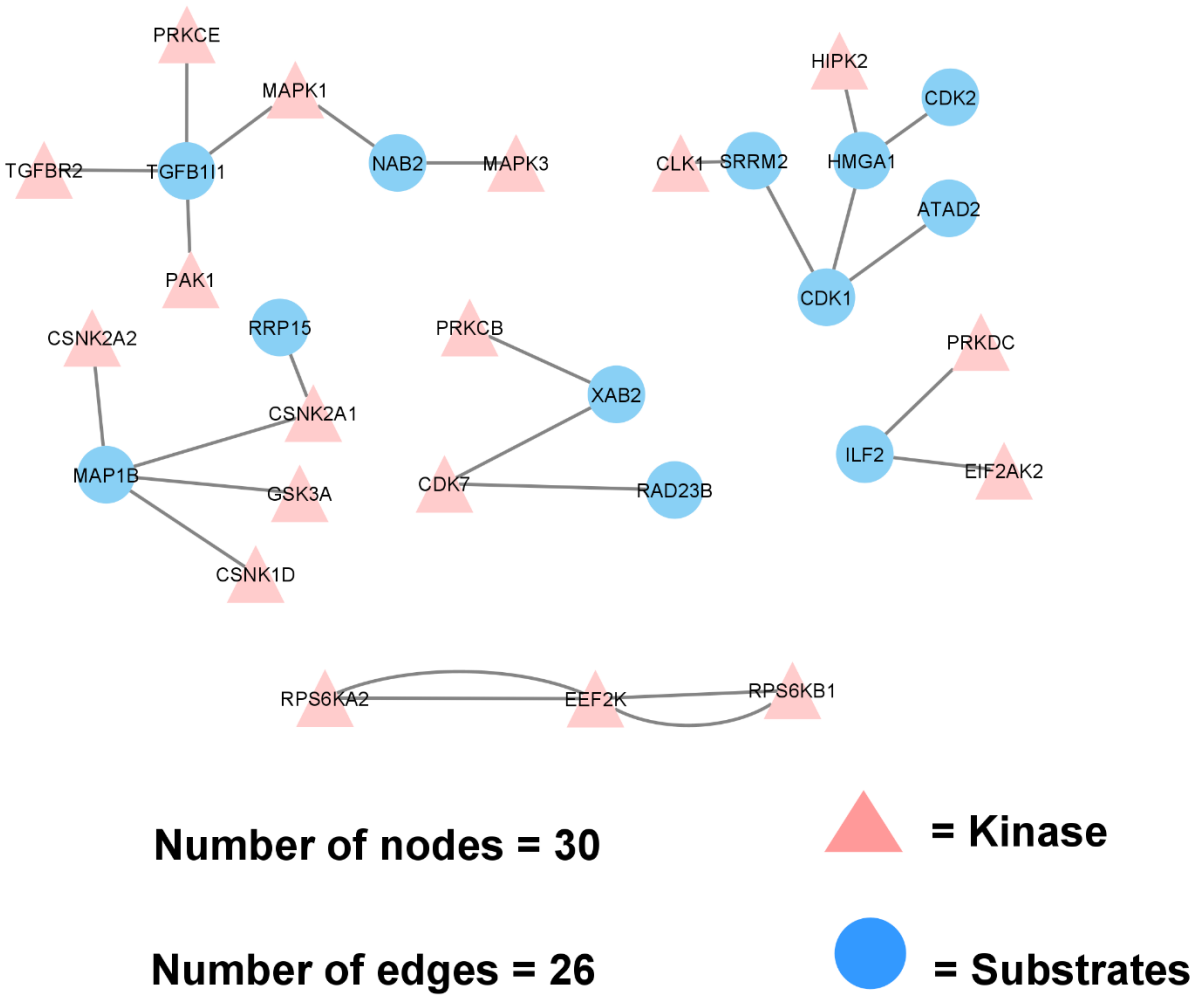


Figure 45: Kinase-Substrate Interaction Network – 15 µM Rucaparib (created with Cytoscape)

CONCLUSION

Breast cancer is a leading cause of death among all other types of cancer in women worldwide. TNBC is the most invasive form of breast cancer and has a poorer prognosis than hormone receptor positive and HER2 type breast cancers. The treatment strategy available for TNBC is chemotherapy combined with PARP inhibitors in BRCA1 and BRCA2-deficient breast cancers. But despite initial success, resistance to PARP1 inhibitors was found in many subtypes of TNBC including off-target effects. Thus, it is important to find alternative treatments that can selectively target cancer cells and also improve sensitivity to PARP1 inhibitors. In this study, TNBC cells overexpressing NQO1 were treated with NQO1 bio activable drug IB-DNQ, which produces ROS accumulating large amounts of H_2O_2 causing DNA damage. Normal cells remove the H_2O_2 using catalase, but most solid cancers including breast cancer cells have low levels of catalase and high expression of NQO1. This difference in NQO1: catalase ratio between normal and cancer cells provides an excellent therapeutic window. The increase in H2O2 levels causes PARP1 hyperactivation which doesn't mean that PARP1 is involved in active DNA repair. On the contrary, PARP1 hyperactivation inhibits the repair activity. PARP1 and PARG are responsible for balance between PAR production and degradation. Any dysregulation of this balance towards overproduction of PAR is detrimental to cells due to NAD+ overconsumption. The PARP1 hyperactivation in response to IB-DNQ induced DNA damage causes huge losses in NAD+ and ATP. The cells die by activating caspase independent programmed necrosis called NAD+ keresis. Addition of PARP1 inhibitor Rucaparib prevents PARP1 hyperactivation and recycles NAD+ back to NADPH which in turn powers more futile redox cycle creating more NQO1 – IB-DNQ mediated DNA

breaks, SSB to DSB conversion and finally death of cancer cells which is consistent with prior work [20]. This shows synergistic tumor selective action between the PARP1 inhibitor and IB-DNQ.

This phosphoproteome study reveals the changes in downstream cellular signaling in response to different concentrations of IB-DNQ and Rucaparib. Large changes in phosphorylation were observed in combination therapy with low dose IB-DNQ (0.1 μM) and Rucaparib (15 μM) compared to the single agent Rucaparib 15 μM, 0.1 μM IB-DNQ and 0.4 μM IB-DNQ. The pathway enrichment using gene ontology (GO) and gene set enrichment analysis (GSEA) revealed expression of proteins related to the cell cycle checkpoint, transcription repression machinery and DNA damage. The protein-protein interaction analysis (PPI) showed a high level of connectivity between proteins related to these pathways. Various computational approaches were used to understand the protein network and clusters of highly interconnected proteins within the PPI.

To elucidate the role of different kinases involved in the phosphorylation events, kinase enrichment analysis was performed using a tool called KSEA. The kinases substrate relationship is based on the published literature from phosphosite specific databases like PhosphoSitePlus and NetworkIN, which uses experimental and computational information, respectively, for annotation. The CDK1 and CDK2 which play important role in DSB repair by homologous recombination by phosphorylation of BRCA1, are found in the downregulated area upon combination drug treatment. This prevents repair of DSB and increases sensitivity to PARP1 inhibitors. Further mapping of the identified kinases from KSEA analysis on a kinome tree reveals upregulation in AGC and CK1 family of kinases.

We also created a network topology map of the kinase substrate relationship to identify highly interconnected proteins or hub proteins. Any perturbation in the hub proteins have very large downstream signaling effects. For example, the CDK1 and CDK2 kinases has many interactions indicating large changes due to IB-DNQ and Rucaparib treatment.

This research is also helpful in understanding protein drug interaction which requires further research related to systems biology and bioinformatics. Overall, the current study has shown that proteomic analysis is a powerful tool to understand the phosphorylation patterns and affected genes upon drug treatment of TNBC cells.

Future Directions

The complexity of signaling networks arises due to myriad of dynamic interactions that flux in time and space. There is a continuum of proteomes which interchange during cellular signaling process and understanding this requires a global view of the signaling networks to identify the systems trajectories that drives these changes. For this purpose, it is increasingly important to use a combination of experimental and computational models to understand the cellular interaction networks at systems levels and how they change the cell fate and behavior over time. The topological properties of the protein-protein and kinase-substrate interaction networks can be studied further to elucidate the directionality and centrality of the networks. The experimental PPI are always undirected which means the directionality of interactions are not provided. Here computational approaches like Silverbush's algorithm [73] or machine learning algorithms can be applied to confer directionality to the edges in the network. Also, proteogenomics approach by integrating

experimental proteomic data, existing PPI information and gene expression data can be used to predict changes indicative of disease progression [74]. Tools like MERLIN [75] and SPAGI [76] can help to identify active signaling pathways by integrating gene expression and protein interaction data.

REFERENCE

[1] Americal Cancer Society, “Cancer facts and figures,” 2019.

[2] R. Kamps *et al.*, “Next-generation sequencing in oncology: Genetic diagnosis, risk prediction and cancer classification,” *Int. J. Mol. Sci.*, vol. 18, no. 2, 2017.

[3] A. J. De Smith, R. G. Walters, P. Froguel, and A. I. Blakemore, “Human genes involved in copy number variation: Mechanisms of origin, functional effects and implications for disease,” *Cytogenet. Genome Res.*, vol. 123, no. 1–4, pp. 17–26, 2009.

[4] M. Beigh, “Next-Generation Sequencing: The Translational Medicine Approach from ‘Bench to Bedside to Population,’” *Medicines*, vol. 3, no. 2, p. 14, 2016.

[5] S. Doll, F. Gnad, and M. Mann, “The Case for Proteomics and Phospho-Proteomics in Personalized Cancer Medicine,” *Proteomics - Clin. Appl.*, vol. 13, no. 2, pp. 1–10, 2019.

[6] I. ur R. Tak, F. Ali, J. S. Dar, A. R. Magray, B. A. Ganai, and M. Z. Chishti, *Posttranslational modifications of proteins and their role in biological processes and associated diseases*, no. January. Elsevier Inc., 2019.

[7] T. M. Karve and A. K. Cheema, “Small Changes Huge Impact: The Role of Protein Posttranslational Modifications in Cellular Homeostasis and Disease,” *J. Amino Acids*, vol. 2011, pp. 1–13, 2011.

[8] B. Ogretmen, “乳鼠心肌提取 HHS Public Access,” *Physiol. Behav.*, vol. 176, no. 3, pp. 139–148, 2019.

[9] G. Giudice and E. Petsalaki, “Proteomics and phosphoproteomics in precision medicine: Applications and challenges,” *Brief. Bioinform.*, vol. 20, no. 3, pp. 767–

777, 2017.

[10] O. A. Arshad *et al.*, "An Integrative Analysis of Tumor Proteomic and Phosphoproteomic Profiles to Examine the Relationships Between Kinase Activity and Phosphorylation," *Mol. Cell. Proteomics*, vol. 18, no. 8 suppl 1, pp. S26–S36, 2019.

[11] J. A. Ubersax and J. E. Ferrell, "Mechanisms of specificity in protein phosphorylation," *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 7, pp. 530–541, 2007.

[12] C. Francavilla *et al.*, "Phosphoproteomics of Primary Cells Reveals Druggable Kinase Signatures in Ovarian Cancer," *Cell Rep.*, vol. 18, no. 13, pp. 3242–3256, 2017.

[13] D. Ross and D. Siegel, "Functions of NQO1 in cellular protection and CoQ10 metabolism and its potential role as a redox sensitive molecular switch," *Front. Physiol.*, vol. 8, no. AUG, pp. 1–10, 2017.

[14] A. K. Jaiswal, "Regulation of genes encoding NAD(P)H:quinone oxidoreductases," *Free Radic. Biol. Med.*, vol. 29, no. 3–4, pp. 254–262, 2000.

[15] A. L. Pey, C. F. Megarity, and D. J. Timson, "NAD(P)H quinone oxidoreductase (NQO1): An enzyme which needs just enough mobility, in just the right places," *Biosci. Rep.*, vol. 39, no. 1, pp. 1–9, 2019.

[16] K. Zhang, D. Chen, K. Ma, X. Wu, H. Hao, and S. Jiang, "NAD(P)H:Quinone Oxidoreductase 1 (NQO1) as a Therapeutic and Diagnostic Target in Cancer," *J. Med. Chem.*, vol. 61, no. 16, pp. 6983–7003, 2018.

[17] E. I. Parkinson, J. S. Bair, M. Cismesia, and P. J. Hergenrother, "Efficient NQO1 substrates are potent and selective anticancer agents," *ACS Chem. Biol.*, vol. 8, no.

10, pp. 2173–2183, 2013.

[18] X. Huang *et al.*, “HHS Public Access,” vol. 72, no. 12, pp. 3038–3047, 2016.

[19] E. I. Parkinson and P. J. Hergenrother, “Deoxynyboquinones as NQO1-Activated Cancer Therapeutics,” *Acc. Chem. Res.*, vol. 48, no. 10, pp. 2715–2723, 2015.

[20] X. Huang *et al.*, “HHS Public Access,” vol. 30, no. 6, pp. 940–952, 2017.

[21] M. Porteous and D. Stirling, “Breast Cancer,” *Brenner’s Encycl. Genet. Second Ed.*, vol. 12, no. 10, pp. 387–390, 2013.

[22] S. Yang *et al.*, “The association between NQO1 Pro187Ser polymorphism and bladder cancer susceptibility: A meta-analysis of 15 studies,” *PLoS One*, vol. 10, no. 1, pp. 1–15, 2015.

[23] S. J. Saldivar *et al.*, “An association between a NQO1 genetic polymorphism and risk of lung cancer,” *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.*, vol. 582, no. 1–2, pp. 71–78, 2005.

[24] O. Moscovitz *et al.*, “A Mutually Inhibitory Feedback Loop between the 20S Proteasome and Its Regulator, NQO1,” *Mol. Cell*, vol. 47, no. 1, pp. 76–86, 2012.

[25] B. Lajin and A. Alachkar, “The NQO1 polymorphism C609T (Pro187Ser) and cancer susceptibility: A comprehensive meta-analysis,” *Br. J. Cancer*, vol. 109, no. 5, pp. 1325–1337, 2013.

[26] L. Cao *et al.*, “Tumor-selective, futile redox cycle-induced bystander effects elicited by NQO1 bioactivatable radiosensitizing drugs in triple-negative breast cancers,” *Antioxidants Redox Signal.*, vol. 21, no. 2, pp. 237–250, 2014.

[27] and R. W. R. Olivia E. Atherton1, Jennifer L. Tackett2, Emilio Ferrer1, “乳鼠心肌提取 HHS Public Access,” *Physiol. Behav.*, vol. 176, no. 1, pp. 139–148, 2018.

[28] M. J. Patterson, Y. Drew, and N. J. Curtin, “Parp,” *Cancer Ther. Targets*, vol. 2–2, no. 301, pp. 913–934, 2017.

[29] M. Ikegami *et al.*, “High-throughput functional evaluation of BRCA2 variants of unknown significance,” *Nat. Commun.*, vol. 11, no. 1, 2020.

[30] G. Yang, C. Sau, W. Lai, J. Cichon, and W. Li, “BRAC1 and BRAC2 mutation and treatment strategies for breast cancer,” *HHS Pulic Access*, vol. 344, no. 6188, pp. 1173–1178, 2015.

[31] E. A. Bey *et al.*, “An NQO1- and PARP-1-mediated cell death pathway induced in non-small-cell lung cancer cells by β-lapachone,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 28, pp. 11832–11837, 2007.

[32] Thermo Fisher, “Proteome Discoverer.” 2019.

[33] N. L. Hervé Pagès, Marc Carlson, Seth Falcon, “Annotationdbi package,” *Bioconductor.org*, 2020. [Online]. Available: https://www.bioconductor.org/packages/release/bioc/html/AnnotationDbi.html.

[34] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, “ClusterProfiler: An R package for comparing biological themes among gene clusters,” *Omi. A J. Integr. Biol.*, vol. 16, no. 5, pp. 284–287, 2012.

[35] G. Yu, “enrichGO,” *Bioconductor.org*, 2020. [Online]. Available: https://rdrr.io/bioc/clusterProfiler/man/enrichGO.html.

[36] G. Yu, “cnetplot,” *Bioconductor.org*, 2020. [Online]. Available: https://rdrr.io/bioc/enrichplot/man/cnetplot.html.

[37] G. Yu, “enrichplot,” *Bioconductor.org*, 2020. [Online]. Available: https://bioconductor.org/packages/devel/bioc/manuals/enrichplot/man/enrichplot.p

df.

[38] A. Subramanian *et al.*, “Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 43, pp. 15545–15550, 2005.

[39] A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, and J. P. Mesirov, “Molecular signatures database (MSigDB) 3.0,” *Bioinformatics*, vol. 27, no. 12, pp. 1739–1740, 2011.

[40] M. Broad Institute, “Morpheus: Versatile matrix visualization and analysis software,” 2005. [Online]. Available: https://software.broadinstitute.org/morpheus/.

[41] D. D. Wiredja, M. Koyutürk, and M. R. Chance, “The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics,” *Bioinformatics*, vol. 33, no. 21, pp. 3489–3491, 2017.

[42] P. V. Hornbeck *et al.*, “PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse,” *Nucleic Acids Res.*, vol. 40, no. D1, pp. 261–270, 2012.

[43] R. Linding *et al.*, “NetworKIN: A resource for exploring cellular phosphorylation networks,” *Nucleic Acids Res.*, vol. 36, no. SUPPL. 1, pp. 695–699, 2008.

[44] Yoav Benjamini and Yosef Hochberg, “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing,” *J. R. Stat. Soc. Ser. B*, vol. 57, no. 1, pp. 289–300, 1995.

[45] O. Wagih, “Ggseqlogo: A versatile R package for drawing sequence logos,” *Bioinformatics*, vol. 33, no. 22, pp. 3645–3647, 2017.

[46] K. S. Metz *et al.*, “Coral: Clear and Customizable Visualization of Human Kinome Data,” *Cell Syst.*, vol. 7, no. 3, pp. 347-350.e1, 2018.

[47] 1 Paul Shannon *et al.*, “Cytoscape: A Software Environment for Integrated Models,” *Genome Res.*, vol. 13, no. 22, p. 426, 1971.

[48] D. Szklarczyk *et al.*, “STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, 2019.

[49] E. P. Rogakou, D. R. Pilch, A. H. Orr, V. S. Ivanova, and W. M. Bonner, “DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139,” *J. Biol. Chem.*, vol. 273, no. 10, pp. 5858–5868, 1998.

[50] P. Mittal and C. W. M. Roberts, “The SWI/SNF complex in cancer — biology, biomarkers and therapy,” *Nat. Rev. Clin. Oncol.*, vol. 17, no. 7, pp. 435–448, 2020.

[51] N. J. N. N. P. S. Daniel E Shumer, “乳鼠心肌提取 HHS Public Access,” *Physiol. Behav.*, vol. 176, no. 12, pp. 139–148, 2017.

[52] J. A. Biegel, T. M. Busse, and B. E. Weissman, “SWI/SNF chromatin remodeling complexes and cancer,” *Am. J. Med. Genet. Part C Semin. Med. Genet.*, vol. 166, no. 3, pp. 350–366, 2014.

[53] S. Linder and P. Kopp, “Podosomes at a glance,” *J. Cell Sci.*, vol. 118, no. 10, pp. 2079–2082, 2005.

[54] D. C. Flynn, Y. Cho, D. Vincent, and J. M. Cunnick, “Podosomes and Invadopodia: Related structures with Common Protein Components that May Promote Breast Cancer Cellular Invasion,” *Breast Cancer Basic Clin. Res.*, vol. 2, p. BCBCR.S789, 2008.

[55] K. D. Pruitt, C. W. Hogue, M. Groll, C. Hartmann, R. Huber, and S. Fields, "Mcode," *Nucleic Acids Res.*, vol. 29, no. 1, pp. 137–140, 2001.

[56] B. Lehne and T. Schlitt, "Protein-protein interaction databases: keeping up with growing interactomes.," *Hum. Genomics*, vol. 3, no. 3, pp. 291–297, 2009.

[57] W. Li *et al.*, "A methylation-phosphorylation switch determines Plk1 kinase activity and function in DNA damage repair," *Sci. Adv.*, vol. 5, no. 3, pp. 1–16, 2019.

[58] Y. J. Jang, S. Ma, Y. Terada, and R. L. Erikson, "Phosphorylation of threonine 210 and the role of serine 137 in the regulation of mammalian polo-like kinase," *J. Biol. Chem.*, vol. 277, no. 46, pp. 44115–44120, 2002.

[59] L. Macůrek *et al.*, "Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery," *Nature*, vol. 455, no. 7209, pp. 119–123, 2008.

[60] K. Strebhardt and A. Ullrich, "for Cancer Therapy," *Cancer*, vol. 6, no. April, pp. 321–330, 2006.

[61] R. E. A. Gutteridge, M. A. Ndiaye, X. Liu, and N. Ahmad, "Plk1 inhibitors in cancer therapy: From laboratory to clinics," *Mol. Cancer Ther.*, vol. 15, no. 7, pp. 1427–1435, 2016.

[62] Z. Liu, Q. Sun, and X. Wang, "PLK1, A potential target for cancer therapy," *Transl. Oncol.*, vol. 10, no. 1, pp. 22–32, 2017.

[63] S. I. King *et al.*, "Immunohistochemical detection of Polo-like kinase-1 (PLK1) in primary breast cancer is associated with TP53 mutation and poor clinical outcom," *Breast Cancer Res.*, vol. 14, no. 2, p. R40, 2012.

[64] Aga Lewelt, "乳鼠心肌提取 HHS Public Access," *Physiol. Behav.*, vol. 176, no. 3, pp. 139–148, 2015.

[65] A. Giordano *et al.*, “Polo-like kinase 1 (Plk1) inhibition synergizes with taxanes in triple negative breast cancer,” *PLoS One*, vol. 14, no. 11, pp. 1–18, 2019.

[66] P. R. Mueller, T. R. Coleman, A. Kumagai, and W. G. Dunphy, “Myt1 : A Membrane-Associated Inhibitory Kinase that Phosphorylates Cdc2 on Both Author (s): Paul R . Mueller , Thomas R . Coleman , Akiko Kumagai and William G . Dunphy Published by : American Association for the Advancement of Science Stable URL : http,” *Science (80-.).*, vol. 270, no. 5233, pp. 86–90, 1995.

[67] A. Petrone, M. E. Adamo, C. Cheng, and A. N. Kettenbach, “Identification of candidate cyclin-dependent kinase 1 (Cdk1) substrates in mitosis by quantitative phosphoproteomics,” *Mol. Cell. Proteomics*, vol. 15, no. 7, pp. 2448–2461, 2016.

[68] J. Bartek and J. Lukas, “Chk1 and Chk2 kinases in checkpoint control and cancer,” *Cancer Cell*, vol. 3, no. 5, pp. 421–429, 2003.

[69] N. Johnson and G. I. Shapiro, “Cyclin-dependent kinases (cdks) and the DNA damage response: Rationale for cdk inhibitorchemotherapy combinations as an anticancer strategy for solid tumors,” *Expert Opin. Ther. Targets*, vol. 14, no. 11, pp. 1199–1212, 2010.

[70] R. I. Yarden and M. Z. Papa, “BRCA1 at the crossroad of multiple cellular pathways: Approaches for therapeutic interventions,” *Mol. Cancer Ther.*, vol. 5, no. 6, pp. 1396–1404, 2006.

[71] L. P. Ferretti, L. Lafranchi, and A. A. Sartori, “Controlling DNA-end resection: A new task for CDKs,” *Front. Genet.*, vol. 4, no. JUN, pp. 1–7, 2013.

[72] A. R. D. levine susan hanson maureen Germain, “乳鼠心肌提取 HHS Public Access,” *Physiol. Behav.*, vol. 176, no. 3, pp. 139–148, 2017.

[73] D. Silverbush and R. Sharan, “A systematic approach to orient the human protein–protein interaction network,” *Nat. Commun.*, vol. 10, no. 1, 2019.

[74] I. W. Taylor *et al.*, “Dynamic modularity in protein interaction networks predicts breast cancer outcome,” *Nat. Biotechnol.*, vol. 27, no. 2, pp. 199–204, 2009.

[75] S. Roy, S. Lagree, Z. Hou, J. A. Thomson, R. Stewart, and A. P. Gasch, “Integrated Module and Gene-Specific Regulatory Inference Implicates Upstream Signaling Networks,” *PLoS Comput. Biol.*, vol. 9, no. 10, 2013.

[76] M. H. Kabir, R. Patrick, J. W. K. Ho, and M. D. O’Connor, “Identification of active signaling pathways by integrating gene expression and protein interaction data,” *BMC Syst. Biol.*, vol. 12, no. Suppl 9, 2018.

CURRICULUM VITAE

Gitanjali Roy

Education

- MS Biochemistry and Molecular Biology earned at Indiana University-Purdue University Indianapolis, January 2021.
- MS Molecular Biology and Genetics earned at Presidency College under University of Calcutta, India, July 2009.
- BS Botany earned at Lady Brabourne College under University of Calcutta, India, July 2007.

Honors and Awards

- Department of Biotechnology, Government of India sponsored studentship in Bioinformatics at Presidency College, Kolkata, India.
- Merit based scholarship from The George Washington University under the 'Global Initiatives Fellowship'.

Research Experience

- Summer project:" In-silico screening of chemical small molecule database against a protein target to develop antibacterial" at The George Washington University, Department of Pharmacology & Physiology under supervision of Dr Nagarajan Pattabiraman.
- 'PI3K signalling pathway and its inhibition as a therapeutic strategy in cancer treatment' at The George Washington University, Department of Pharmacology & Physiology under supervision of Dr. Norman H Lee.

Professional Experience

- 9 years' experience as Executive- QA/QC at Dabur India Limited, India.

Conferences Attended

- Attended HUPO connect 2020 conference.
- Poster presentation on "Phosphoproteome Analysis Of Metastatic Breast Cancer Cells Following Treatment With a Tumor-selective NQO1 Bioactivatable Drug" Page 52 at the US HUPO 16th Annual Conference 2020.
- Abstract publication and Oral Presentation titled "Physiological and Biochemical alterations of cold stressed barley seeds (*Hordeum vulgare L*.) and it's allelopathic effect" at 36th Annual conference of "Ethological Society of India" November 26-28, 2012, Kolkata, India, p. 34.